Simple principal components.

Thesis

Simple Principal Components

A thesis submitted to the Open University for the degree of Doctor of Philosophy

Linjuan Sun
The Open University

December 2005

DATE OF SUBMISSION: 22 DECEMBER 2005
DATE OF AWARD: 4 MAY 2006

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

For my husband Ruijiu Nie

Without whose encouragement I would never have finished this thesis

Acknowledgments

Thanks are due to a number of people who have helped me:

Dr Karen Vines, she read the draft of the thesis many many times, gave me a lot of significant suggestions, and also offered a lot of help in my research.

Professor Frank Critchley read the draft of my thesis twice, gave me lots of very good advice.

Dr Catriona Queen often discuss my progress on my Ph.D.

This thesis would not be possible without their help.

Abstract

A new technique, simple principal component analysis (SCA), is addressed by Vines (2000) to enhance the interpretation of principal components (PCs). The SCA algorithm seeks integer valued loadings vectors that have properties close to the loading vectors obtained from the principal component analysis (PCA). Simulation is used to compare the different implementations and show that SCA is better than PCA in some cases. In this thesis, I first pin down the link between SCA with Jacobi methods, then develop the concepts of a combining approach and a hybrid approach.

The results produced by SCA methods can be very good approximation to corresponding PCs whatever the structures of the data, and simple components are generally easier to interpret than PCs. In particular, the sample simulation results of SCA are generally better than PCA for any simple structures. SCA2, SCA5 and SCA6 are the best SCA methods. As the dimension of the data increases, SCA5 and SCA6 are similar. If necessary, combined and hybrid approaches can make the results of SCA more accurate. However, the results of hybrid methods are orthogonal but the combined results are generally not orthogonal.

Contents

List of Tables

List of Figures

Chapter 1

Introduction-Principal components

1.1 Introduction

Principal component analysis (PCA) is a widely used statistical technique (Jolliffe, 2002b) in multivariate statistical analysis. The main aim of PCA is to reduce a large number of variables (say p) to a smaller number (say m) of principal components (PCs) while retaining as much as possible of the variation in the original variables. This chapter describes the concept of principal components, along with four examples of applications of PCA. Section 1.2 gives a brief description of PCA and of the notation which will be used in this thesis. Section 1.3 gives some examples of application of PCA. Section 1.4 given a discussion and some conclusions for this chapter. Section 1.5 gives an outline of this thesis.

1.2 Brief description of PCA

Principal component analysis is a dimension-reduction technique that is applied to multivariate data. The aim of PCA is to represent p (possibly) interrelated variables in terms of a smaller number of uncorrelated components (say m). These m uncorrelated variables are defined as the principal components (PCs), and they are linear combinations of the original variables. The principal components are defined as follows.

Suppose x be a vector of p random variables, x_1, x_2, ..., x_p, i.e. $x = (x_1, x_2, \ldots, x_p)$ with variance covariance matrix Σ. The k^{th} principal component z_k, for $k = 1, 2, \ldots, m$, is defined as the linear combination $x\alpha_k$, $\alpha_k = (\alpha_{k1}, \alpha_{k2}, \ldots, \alpha_{kp})^T$, which maximizes $var[x\alpha_k] = \alpha_k^T \Sigma \alpha_k$ under the conditions $\alpha_k^T \alpha_k = 1$, and (for $k \geq 2$), $\alpha_j^T \alpha_k = 0$, $j \leq k$. $z = (z_1, \ldots, z_m)$.

In general, the k^{th} PC of x in the previous paragraph is $x\alpha_k$ and $var(x\alpha_k) = \lambda_k$, where λ_k is the k^{th} largest eigenvalue of Σ, and α_k is the corresponding eigenvector. α_k is termed the loadings vector for the k^{th} PC. The matrix $A = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ is called the loading matrix. So the PCs are defined as $z = xA$. Thus, the PCs are defined by an orthonormal linear transformation of x. Also $var[x\alpha_k] = \alpha_k^T \Sigma \alpha_k = \lambda_k$, $k = 1, 2, \ldots, n$, i.e. $A^T \Sigma A = \Lambda$, where Λ is the diagonal matrix whose k^{th} diagonal element is λ_k. So the PCs are uncorrelated. This property is nice because it implies that the variance explained by one principal component is unaffected by that of another principal component. So the percentage of variance explained by the k^{th} PC is equal to $\lambda_k / \sum_{i=1}^{p} \lambda_i$. It can be shown that principal components are optimal under different criteria for orthogonal or uncorrelated components (Rao 1964; Okamoto 1969; McCabe 1984).

The m principal components are ordered by the relative percentages of total variation accounted for by each component, the first component accounting for the biggest percentage of variation, then the second the next biggest percentage of variation and so on. So the first few components will represent the most variation that can be captured in any m components.

The major aim of PCA is to replace the p variables by a much smaller number (m) of PCs, which nevertheless lose little variation. So it is necessary to decide on the number of principal components to be retained without much information loss. A simple way based on using a correlation matrix is to consider only those components which have eigenvalues of 1.00 or greater. Another popular way is to select the m principal components which make the cumulative percentage of total variance exceed some percentage (say 70%). This is the way used in the examples of this chapter. There are several other formal ways to determine m, such as cross-validation and bootstrapping. These methods will not be discussed here because they are not used in this thesis.

In order to distinguish the results of PCA from any method used to modify PCs, the k^{th} component of a linear combination of the original variables x (which is not a PC) is defined as

$w_k = x\gamma_k$, $k = 1, 2, \ldots, p$, $\gamma_k = (\gamma_{k1}, \ldots, \gamma_{kp})^T$. Suppose $w = (w_1, w_2, \ldots, w_p)$, then C is the loading matrix of a linear combination, i.e. the k^{th} column of C is γ_k, thus $w = xC$.

In practice, PCA generally starts with a $n \times p$ data matrix X, where n and p are the number of observations and the number of variables respectively. The i^{th} row of X, X_i, is the value of vector x corresponding to the i^{th} observation. Thus the value of the k^{th} PC for the i^{th} observation is represented by $X_i\alpha_k$ $(i = 1, \ldots, n)$ in the sample case. So the k^{th} PC corresponding to n observations are $Z_k = X\alpha_k$, $k = 1, 2, \ldots, m$. If $Z = (Z_1, Z_2, \ldots, Z_m)$, the matrix of PC scores is $Z = XA$. Similarly, a linear combination by any other method is $W_k = X\gamma_k$, $W = (W_1, \ldots, W_m)$. The matrix of component scores is $W = XC$. When the column means of X are 0, the sample covariance matrix of X can be written as $S = \frac{1}{n-1}X^TX$. Note that the eigenvectors matrices of $\frac{1}{n-1}X^TX$ and X^TX are identical. So it will be convenient for some of the approaches to work in terms of eigenvalues and eigenvectors of X^TX, rather than with those of S. PCs has two nice properties, orthogonal and uncorrelated. PCA is often done for variables standardized to each have unit variance, so Σ or S are the same as the correlation matrix for the original variables.

PCA can be done via the singular value decomposition (SVD) of the data matrix (Section 3.5, Jolliffe, 2002b). For the SVD of X, $X = ULA^T$ where U is a $n \times r$ matrix whose k^{th} column is the k^{th} PC with unit length, A is the $n \times r$ loadings matrix of the principal components and r is the rank of X. The variance of the k^{th} PC is $L_{kk}^2/(n-1)$, L is an $r \times r$ diagonal matrix (in Jolliffe, 2002b, the SVD start from X^TX, L_{kk}^2 is the eigenvalues of X^TX rather than S). So the matrix of PC scores is $Z = XA = ULA^TA = UL$, or $U = ZL^{-1}$.

1.3 Examples

There have been many applications of PCA. Here are a few examples.

1.3.1 RI data

First is an application of PCA to Resistance Index (RI) data. According to Vines (2000), "RI is a measure of resistance to flow in blood-vessels. A large RI indicates that a blood-vessel is

Variables	1	2	3	4
Dop-R	1.55	0.70	1.26	0.67
Dop-L	0.70	1.53	0.63	1.30
Cvi-R	1.26	0.63	2.54	0.76
Cvi-L	0.67	1.30	0.76	2.66

Table 1.1: The covariance matrix of 4 measures for the RI data

		Components			
		1	2	3	4
Variables	Dop-R	0.42	0.31	0.56	0.62
	Dop-L	0.43	-0.30	0.55	-0.65
	Cvi-R	0.55	0.65	-0.43	-0.30
	Cvi-L	0.58	-0.63	-0.42	0.31
	Variance	4.81	2.15	0.79	0.54
	Cumulative Variance	58%	84%	94%	100%

Table 1.2: Principal components for the RI data

more resistant to blood flow. In a study on the ultrasound monitoring of pregnant woman, RI measurements were taken on 444 women scanned between 18 and 32 weeks' gestation (Thompson et al., 1999). On each woman four separate measurements of the RI from the uterine artery were recorded. These corresponded to the RI in the uterine artery on each side of the body (right and left) using two different techniques (Doppler and colour velocity imaging (CVI)).” So there are 4 variables in this example, right-side Doppler RI, left-side Doppler RI, right-side CVI RI and left-side CVI RI respectively. The variance-covariance matrix (multiplied by 100) of these data is given in Table1.1.

Because all measurements are on the same scale, it is reasonable to use the variance-covariance matrix. Performing PCA on RI data gives the results shown in Table 1.2.

All the loadings of the first component are very similar and have the same sign, thus the first component represents an average of all four measurements. This component represents 58% of the variance. The second component represents the difference between the RI in the uterine arteries on the right and left sides of the body. Furthermore the difference in the uterine arteries on the right and left sides of the body for the CVI technique is more important (about twice as much) than that for the Doppler technique. This component represents 26% of the variance. The third component gives a difference between measurements of the RI using the two different techniques. The component represents 10% of the variance. The fourth component denotes the interaction between location and technique, the difference in the uterine arteries on the right and left sides of the body for Doppler technique being more important than the difference in the uterine arteries on the left and right sides of the body for CVI technique. The component represents 6% of the variance. The first two components of PCA account for 84% of the variance of RI data, so it is enough just to consider the first two components for this data, i.e $m = 2$. Even for this simple example, the exact loadings of the PCs make the interpretation of components 2 and 4 not very clear because the loadings of these two components are not very big (near 1) or very small (near zero) and the loadings lie in a long range.

1.3.2 Sparrow data

The second example is an analysis of five body measurements of female sparrows (data originally given by Bumpus (1898) and reanalysed by Manly (1995)). After a severe storm on 1898, sparrows were taken to the biological laboratory at Brown University. A lot of morphological measurements on each bird were recorded. Manly (1995) considered 5 variables. Variable 1 (measurement 1) was the total length of the bird, variable 2 was the alar extent of the bird, variable 3 was the length of beak and head of the bird, variable 4 was the length of humerus of the bird and variable 5 is the length of keel of the sternum of the bird. For this example, if the covariance matrix is used, the variances of the individual variables are widely different (the biggest variance is about 83 times of the smallest variance). So the correlation matrix (Table 1.3) between the five body measurements of female sparrows is used.

Performing PCA on the sparrow data gives the results in Table 1.4. Clearly the first com-

Variables	1	2	3	4	5
total length	1.00	0.74	0.66	0.65	0.61
alar extent	0.74	1.00	0.67	0.77	0.53
length of beak and head	0.66	0.67	1.00	0.76	0.53
length of humerus	0.65	0.77	0.76	1.00	0.61
length of keel of sternum	0.61	0.53	0.53	0.61	1.00

Table 1.3: Correlations between the five body measurements of female sparrows

ponent is far more important than the others, because it accounts for 72% of the variance. All the loadings of the first principal component are very similar and have the same sign, so the first component is an average of all the five measurements. So this is an index of the size of the sparrows. The second principal component appears to be a contrast between the alar extent, length of beak and head, and length of humerus on the one hand, and the length of the keel of the sternum on the other. The length of the keel of the sternum plays an important role. Total length can be ignored because its loading is too small. However the second component is not easy to interpret because of the exact loadings and because the loadings lie in wide range. The third component represents the difference between total length and alar extent on one side, and length of beak and head, length of humerus on the other side. Clearly each variable does not have the same importance. The loadings make the interpretation of this PC hard because the loadings are not very similar or not very big or not very small. The fourth component is the contrast between total length and length of beak and head on one side, and alar extent and length of humerus on the other. This interpretation ignores the last variable, length of keel of sternum, and considers the other variables are of equal importance. The last component appears to be the difference between total length and length of humerus on one side, and alar extent, length of beak and head and length of keel of sternum on the other. The interpretation of this component are not very clear about which variable is more important because of the exact loadings of the component. So components 2 to 5 represent some aspects of shape differences. This makes sense because component 1 captures size only .

	Components				
Variables	1	2	3	4	5
total length	0.45	0.05	-0.69	-0.42	-0.37
alar extent	0.46	-0.30	-0.34	0.55	0.53
length of beak and head	0.45	-0.33	0.45	-0.61	0.34
length of humerus	0.47	-0.18	0.41	0.39	-0.65
length of keel of sternum	0.40	0.88	0.18	0.07	0.19
Variance	3.62	0.53	0.39	0.30	0.16
Cumulative Variance	72%	83%	91%	97%	100%

Table 1.4: Principal components for the sparrow data

The five components explain 72%, 11%, 8%, 6% and 3% variance respectively. The first two components of PCA account for 83% of the variance of sparrow data, so it is enough just to consider the first two components (i.e. $m = 2$) for this data. Except for component 1, the exact loadings of the other components make interpretation more difficult even if the small loadings are ignored and rounded loadings are used.

1.3.3 Employment in Europe data

The third example is some data on employment in European countries (data originally from Euromonitor (1979) and reanalysed by Manly (1995)). This dataset gives the percentages of the labour force in nine different types of industry for 26 European countries. Nine variables are included in the analysis each corresponding to a different industry type, X_1-Agriculture, X_2-Mining, X_3-Manufacturing, X_4-Power supplies, X_5-Construction, X_6-Service industries, X_7-Finance, X_8-Social and personal services, X_9-Transport and communications. The correlation matrix for these data is given in Table 1.5.

Performing PCA on employment in Europe data gives the result in Table 1.6. The first 3 components explain 39%, 24% and 12% variance respectively. That is to say, the first three

Variables	1	2	3	4	5	6	7	8	9
1	1.00	0.04	-0.67	-0.40	-0.54	-0.74	-0.22	-0.75	-0.57
2	0.04	1.00	0.45	0.41	-0.03	-0.40	-0.44	-0.28	0.16
3	-0.67	0.45	1.00	0.39	0.50	0.20	-0.16	0.15	0.35
4	-0.40	0.41	0.39	1.00	0.06	0.20	0.11	0.13	0.38
5	-0.54	-0.03	0.50	0.06	1.00	0.36	0.02	0.16	0.39
6	-0.74	-0.40	0.20	0.20	0.36	1.00	0.37	0.57	0.19
7	-0.22	-0.44	-0.16	0.11	0.02	0.37	1.00	0.11	-0.25
8	-0.75	-0.28	0.15	0.13	0.16	0.57	0.11	1.00	0.57
9	-0.57	0.16	0.35	0.38	0.39	0.19	-0.25	0.57	1.00

Table 1.5: The correlation matrix for percentages employed in nine industry groups in 26 countries in Europe

components of PCA account for 75% of the variance in the data, so it is enough just to retain the first 3 components (i.e. $m = 3$). The first principal component is a difference between agriculture because of a high negative loading and all the other groups of industries except mining and finance. Mining and finance are excluded because the loadings of these two components are very small. This component therefore measures the extent to which people are employed in industries rather than agriculture. The second principal component appears to be a difference between mining, manufacturing, power supplies and transport and communications on one side and service industries, finance and social and personal services on the other side, mining plays a main role because of a high positive loading. So the second component measures the extent to which people are employed in production industries rather than service areas. Agriculture and construction are ignored because of smaller loadings. The third component represents the difference between construction, social and personal services and transport and communications on one side, and mining, manufacturing, power supplies, services industries and finance on the other side. So the third component measures the extent to which people are employed in power

Variables	Components 1	2	3
1-Agriculture	-0.52	0.05	-0.05
2-Mining	-0.00	0.62	0.20
3-Manufacturing	0.35	0.36	0.15
4-Power supplies	0.26	0.26	0.56
5-Construction	0.33	0.05	-0.16
6-Service industries	0.38	-0.35	0.12
7-Finance	0.07	-0.45	0.59
8-Social and personal services	0.39	-0.22	-0.31
9-Transport and communications	0.37	0.20	-0.37
Variance	3.49	2.13	1.10
Cumulative variance	39%	62%	75%

Table 1.6: Principal components for the employment in Europe data

Variables	Components Definition
1-Topdiam	the top diameter of the prop in inches
2-Length	the length of the prop in inches
3-Moist	the moisture content of the prop, expressed as a percentage of the dry weight
4-Testsg	the specific gravity of the timber at the time of test
5-Ovensg	the oven-dry specific gravity of the timber
6-Ringtop	the number of annual rings at the top of the prop
7-Ringbot	the number of annual rings at the base of the prop
8-Bowmax	the maximum bow in inches
9-Bowdist	the distance of the point of maximum bow from the top of the prop in inches
10-Whorls	the number of knot whorls
11-Clear	the length of clear prop from the top of the prop in inches
12-Knots	the average number of knots per whorls
13-Diaknot	the average diameter of the knots in inches

Table 1.7: Definition of variables in the pitprop data

supplies and finance rather than social and personal services and transport and communication. So with this example it starts getting difficult to interpret the PCs, even if the small loadings are ignored.

1.3.4 Jeffers' pitprop data

The last data set considered in this section is the Jeffers' pitprop data set (Jeffers (1967)). There are 13 variables, the definition of the variables are given in Table 1.7.

The correlation matrix of Jeffers' pitprop data is given in Table 1.8. Performing PCA on pitprop data gives the results in Table 1.9. The first 4 components explain 33%, 18%, 14% and 9% variance respectively. So the first four components of PCA account for 74% of the variance

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	0.95	0.36	0.34	-0.13	0.31	0.50	0.42	0.59	0.55	0.08	-0.02	0.13
2	0.95	1	0.30	0.28	-0.12	0.29	0.50	0.42	0.65	0.57	0.08	-0.04	0.14
3	0.36	0.30	1	0.88	-0.15	0.15	-0.03	-0.05	0.13	-0.08	0.16	0.22	0.13
4	0.34	0.28	0.88	1	0.22	0.38	0.17	-0.06	0.14	-0.01	0.10	0.17	0.02
5	-0.13	-0.12	-0.15	0.22	1	0.36	0.30	0.00	-0.04	0.04	-0.09	-0.15	-0.21
6	0.31	0.29	0.15	0.38	0.36	1	0.81	0.09	0.21	0.27	-0.04	0.02	-0.33
7	0.50	0.50	-0.03	0.17	0.30	0.81	1	0.37	0.47	0.68	-0.11	-0.23	-0.42
8	0.42	0.42	-0.05	-0.06	0.00	0.09	0.37	1	0.48	0.56	0.06	-0.36	-0.20
9	0.59	0.65	0.13	0.14	-0.04	0.21	0.47	0.48	1	0.53	0.09	-0.13	-0.08
10	0.55	0.57	-0.08	-0.01	0.04	0.27	0.68	0.56	0.53	1	-0.32	-0.37	-0.29
11	0.08	0.08	0.16	0.10	-0.09	-0.04	-0.11	0.06	0.09	-0.32	1	0.03	0.01
12	-0.02	-0.04	0.22	0.17	-0.15	0.02	-0.23	-0.36	-0.13	-0.37	0.03	1	0.18
13	0.13	0.14	0.13	0.02	-0.21	-0.33	-0.42	-0.20	-0.08	-0.29	0.01	0.18	1

Table 1.8: The correlation matrix of the pitprop data

Variables	Components			
	1	2	3	4
1-Topdiam	-0.40	0.22	-0.21	-0.09
2-Length	-0.41	0.19	-0.24	-0.10
3-Moist	-0.12	0.54	0.14	0.08
4-Testsg	-0.17	0.46	0.35	0.06
5-Ovensg	-0.06	-0.17	0.48	0.05
6-Ringtop	-0.28	-0.01	0.48	-0.06
7-Ringbot	-0.40	-0.19	0.25	-0.07
8-Bowmax	-0.29	-0.19	-0.24	0.29
9-Bowdist	-0.36	0.02	-0.21	0.10
10-Whorls	-0.38	-0.25	-0.12	-0.21
11-Clear	0.01	0.21	-0.07	0.80
12-Knots	0.12	0.34	0.09	-0.30
13-Diaknot	0.11	0.31	-0.33	-0.30
Variance	4.22	2.38	1.88	1.11
Cumulative variance	33%	51%	65%	74%

Table 1.9: Principal components of the Pitprop data

of the pitprop data. Thus it is enough just to consider the first 4 components (i.e. $m = 4$). The first component is a difference between the first 10 variables except the oven-dry specific gravity of the timber (loading is too small to ignore) on one side and the last 3 variables except the length of clear prop because loading is too small to ignore on the other side. The second principal component appears to be a contrast between the top diameter of the prop, the length of the prop, the moisture content of the prop, the specific gravity of the timber, the length of clear prop from the top of the prop, the average number of knots per whorls, the average diameter of the knots and the oven-dry specific gravity of the timber, the number of annual rings at the base of the prop, the maximum bow, the number of knot whorls. The third component represents the difference between the top diameter of the prop, the length of the prop, the maximum bow, the distance of the point of maximum bow from the top of the prop, the number of knot whorls and the average diameter of the knots per whorls on one side, and the moisture content of the prop, the specific gravity of the timber, the oven-dry specific gravity of the timber, the number of annual rings at the top and at the base of the prop on the other side. The fourth component is the contrast between the length of the prop, the number of knot whorls, the average number of knots per whorls and the average diameter of the knots on one side, and the maximum bow and the length of clear prop from the top of the prop on the other. It's very difficult to have a clear interpretation for the first 4 PCs even after ignoring the small loadings of the components.

1.4 Conclusion and discussion

The interpretation of principal components is usually neither easy nor straightforward (e.g. Jeffers 1967; Jackson 1991; Cadima and Jolliffe 1995; Rousson and Gasser 2004). PCs are easier to interpret if the associated loadings are clear-cut, with as few as possible large (in absolute value) loadings (near 1) and as many as possible small (in absolute value) loadings (near 0) or the absolute values of the loadings are very similar. All the examples in this chapter show that if the exact loadings of principal components are not very large and not very small and lie in a wide range, the interpretation of corresponding PCs are difficult, even if the variables with small loadings are ignored and loadings are rounded. In particular the employment in Europe data

(Section 1.3.3) and Jeffers' pitprop data (Section 1.3.4) are difficult to interpret because some of the PCs in these two examples have few large loadings and very few small loadings. Most of the loadings are intermediate in value, that is, the absolute values of the loadings are not very large and not very small and lie in a wide range. For example, consider the third PC of employment in Europe data (Section 1.3.3). This PC has large loadings on variables representing power supplies and finance, a small loading on the variable representing agriculture, and intermediate valued loadings on the other 6 variables.

Various kinds of approaches have been pursued for the simplification and interpretation of principal components. These approaches are reviewed in next chapter.

1.5 Outline of the rest of the thesis

This thesis consists of twelve chapters. Chapter 2 introduces various approaches to simplify the principal components in order to enhance the interpretation of PCs. Chapter 3 considers the simple components algorithm (SCA) provided by Vines (2000). I first pin down the link between the SCA algorithm and the Jacobi algorithm (an iterative algorithm, the algorithm works by repeated application of orthogonal Jacobi rotations, starting from a symmetric matrix), and so explaining why the SCA algorithm is called a Jacobi-like algorithm. Also six different SCA methods are introduced.

In Chapters 4 to 7, population results and simulation results are discussed, the results of PCA are compared with those of different SCA methods. Chapter 4 discusses the results of SCA methods which are applied to the variance-covariance matrix V_0 for 6 dimensional data, the results applied to V_0 are called population results. Chapter 5 investigates sample simulation results for the same 6 dimensional data sets as in Chapter 4. In each simulation, the simulation results based on 500 data sets, each data set having 500 observations. Each data set will be assumed to have a normal distribution $N(0, V_0)$. Chapter 6 discusses the population results for large dimensional data sets. Chapter 7 discusses the sample simulation results for large dimensional data sets. Chapter 8 gives further population results. This chapter includes the results of SCA methods using different k, where k is a tuning parameter which determines the

number of directions considered for each simplicity-preserving transformation. This chapter also investigate the population results based on the same eigenvectors matrix but different eigenvalues.

In order to get simple components, the SCA methods under some restrictions of the directions available at each iteration must be used. Chapter 9 compares the results obtained by SCA methods which use different restrictions on the directions available at each iteration. It will be shown generally that different SCA methods or even the same SCA method using different restrictions, may obtain different results for the same data set. So in Chapter 10, an idea about the combined results is considered. The combined results are obtained by combining the best components from different SCA methods and different restrictions. Some examples and ways to measure the variance interpreted by combined results are given. Chapter 11 discusses a hybrid approach, that is using different SCA methods in different iterations. Finally, the last chapter discusses and outlines further work which could be done.

Chapter 2

Approaches to simplify PCs

2.1 Introduction

It was shown in Chapter 1 that in general it is difficult to achieve a satisfactory interpretation for principal components. In order to enhance interpretation of PCs, the basic PCA technique has been extended or modified by many approaches. In this chapter, these approaches are reviewed. Following Jolliffe (2002b), these techniques can be divided into two groups according to whether or not "simple components" are obtained directly. One group consists of two stage techniques. These techniques first obtain the PCs and then do something with them (Jolliffe, 2002a, 2002b). For example, two stage techniques rotate PCs or replace smaller loadings of PCs by zero. The other group obtains some sort of "simple components" directly.

Detailed description of the approaches starts with two stage approaches. One popular approach is to rotate the principal components, i.e. to first obtain PCs, then rotate to them. Section 2.2 talks about orthogonal and oblique rotation approaches. Section 2.3 deals with truncation of loadings, i.e. first obtain PCs, then truncate the loadings. Section 2.4 discusses methods for generating discrete valued loadings. All but one of these techniques restrict the loadings of the components to integers, whilst the other drives some of the loadings to 0. Most of the techniques are one stage approaches, but one is a two stage approach. Section 2.5 reviews least absolute shrinkage and selection operator approaches applied to PCA. The one stage ap-

proach SCoTLASS and the two stage approach SPCA are discussed. In particular, Section 2.6 talks about other one stage approaches where PCA followed by rotation is replaced by one stage combining variance maximization and simplicity. Section 2.6 discusses the one stage approaches SCoT and SCA_{RG}. The last section, Section 2.7, gives the discussion and conclusion.

2.2 Rotation

Rotation aims to get simple structure by rotating the loadings matrix of principal components. There are two different types of rotation which can be carried out, orthogonal and oblique. Orthogonal rotation refers to the procedure where the columns of the rotation matrix are orthogonal with each other. In contrast to orthogonal rotation, the columns of a matrix for oblique rotation are not orthogonal. Oblique rotation allows for some correlation between the components.

2.2.1 Orthogonal methods

Suppose that we have decided to retain and rotate m PCs. Orthomax rotations start with the general expression

$$S_0(B) = \sum_{j=1}^{m}[\sum_{i=1}^{p} {b_{ij}}^4 - \frac{c}{p}(\sum_{i=1}^{p} {b_{ij}}^2)^2]. \tag{2.1}$$

where B is the rotated matrix of PC loadings matrix, b_{ij} is the $(i,j)^{th}$ element of B, c is an arbitrary constant, determined by the method, and p is the number of variables. In other words, $B = AR$, where, as in Section 1.2, A is the $p \times m$ matrix of PC loadings and R is a $m \times m$ orthogonal matrix. R is obtained so that $S(B)$ is maximized. $S(B)$ is a simplicity criterion. The larger $S_0(B)$ is, the simpler the rotated components. A popular orthogonal rotation method is the varimax method (Kaiser, 1958), which is obtained by maximizing $S_0(B)$ with $c = 1$. The idea behind the varimax criterion is to simplify the structure of the loadings by maximizing the variance of squared loadings within each column of B. This drives the loadings towards 0, 1 or -1. The idea is that each variable should be either very important or very unimportant in a rotated component, with as few cases as possible of borderline importance. There are several

Variables	Components 1	2
Right-side Doppler RI	0.09	-0.52
Left-side Doppler RI	0.51	-0.08
Right-side CVI RI	-0.05	-0.85
Left-side CVI RI	0.85	0.05
Angle	44°	44°
Variance	3.54	3.42
Cumulative variance	43%	84%

Table 2.1: Orthogonal rotated PCs for the RI data

other methods for orthogonal rotation such as the quartimax ($c = 0$) rotation, and the equimax ($c = m/2$) rotation. However Kaiser's varimax method is the most widely accepted method for analytical rotation.

Rotation takes place within the subspace defined by the first m PCs. So that the total variation in this subspace is preserved by the rotation but it is redistributed among the rotated PCs. If $m = p$, then the rotation just changes the order of the original variables.

The first example given here relates to the results from PCA applied to RI data. Recall that for the RI data $m = 2$ (Section 1.3.1). The varimax rotation result of RI data is given in Table 2.1, where the angle is the angle between the original component and rotated component.

After the rotation, the simplicity criterion $S_0(B)$ for RI data is 0.69. This is larger than that of the original PCs ($S_0(B) = 0.12$). So the rotated components are more simple than unrotated PCs. The interpretation of this rotated components is now straightforward. The first component represents measurement on the left side of the body, the second component represents the RI in uterine arteries on the right of the body. However, the variance of the first component decreases by 15% from 58% to 43% and the variance of the second component increases by 15% from 26% to 41%. The cumulative variance by these two rotated components are the same (84%) as that accounted for by the unrotated components. That's what should happen. All the variation in

Variables	Components 1	2
Total length	0.37	0.26
Alar extent	0.55	-0.05
Length of beak and head	0.55	-0.07
Length of humerus	0.50	0.06
Length of keel of sternum	-0.06	0.96
Angle	28°	28°
Variance	2.93	1.22
Cumulative variance	59%	83%

Table 2.2: Orthogonal rotated PCs for the sparrow data

the subspace is preserved by the rotation but it is redistributed amongst the rotated PCs. The angles between the original and rotated components are the same 44°. The angles should be the same for $m = 2$ after orthogonal rotation. This angle is equal to the angle the axes rotated.

The second example is from PCA applied to the sparrow data (Section 1.3.2). Recall that again $m = 2$. The varimax rotation result of sparrow data is listed in Table 2.2. The values of the simplicity criteria of rotated components and original PCs are $S_0(B) = 0.72$ and $S_0(B) = 0.42$ respectively, so the rotated components are more simple than original PCs. The interpretation of these rotated vectors is also straightforward. The first component represents the average of all the variables except the length of keel of sternum. The second component represents the length of keel of sternum. Thus the rotation makes the components easier to interpret compared with the PCs in Section 1.3.2. This is especially true for the second component because only one variable, the last variable, is very important, and three variables (variables 2, 3 and 4) can be ignored because of their small loadings. Comparing the variance of rotated PCs with the original PCs, the variance explained by the first component decreases 13% (from 72% to 59%), the variance explained by the second components increases 13% (from 11% to 24%). As with the RI data, again the total variation is kept, the first two components still explain 83% of the

Variables	Components 1	2	3
1-Agriculture	-0.46	-0.24	-0.12
2-Mining	-0.27	0.54	-0.23
3-Manufacturing	0.14	0.49	-0.08
4-Power supplies	-0.07	0.60	0.29
5-Construction	0.33	0.11	-0.12
6-Service industries	0.40	-0.01	0.35
7-Finance	-0.01	0.02	0.75
8-Social and personal services	0.53	-0.13	-0.06
9-Transport and communications	0.40	0.12	-0.38
Angle	29°	45°	40°
Variance	3.02	2.15	1.54
Cumulative variance	34%	58%	75%

Table 2.3: Orthogonal rotated PCs for the employment in Europe data

variance. Again the angles between the original and rotated components (28°) are the same for the two components.

Finally, consider the application of varimax to the PCA results for the Employment in Europe data described in Section 1.3.3. For this data $m = 3$, the varimax rotation result is given in Table 2.3. The simplicity criterion of rotated components is $S_0(B) = 0.49$, and of original components is $S_0(B) = 0.31$. So the interpretation improves. The rotated first component is the difference between the agriculture, mining and construction, service industries, social and personal services, transport and communications. The second rotated component represents the difference between agriculture and mining, manufacturing, power supplies. The third rotated component denotes the difference between mining, transport and communications and power supplies, service industries, finance. Here the orthogonal rotation is not successful enough to

minimize the number of variables which have a high loading on a component for this example. The variance for the first component decreases (3.49 vs 3.02), the variances for the second and third components increase (2.13 vs 2.15, and 1.10 vs 1.54). Thus again, the total variation is kept but the variance has been redistributed among the rotated PCs. In this example, the angles between the original and rotated components are different with each other. When $m = 3$, this is possible for orthogonal rotation.

So, for these three examples, the orthogonal rotation has been successful for the RI data and the sparrow data but not really successful for the employment in Europe data.

2.2.2 Oblique methods

For oblique rotation, the simplicity criterion $S(B)$ becomes

$$S(B) = \sum_{g<j=1}^{m(m-1)/2} [p \sum_{i=1}^{p} b_{ij}^2 b_{ig}^2 - c(\sum_{i=1}^{p} b_{ij}^2)(\sum_{i=1}^{p} b_{ig}^2)]. \tag{2.2}$$

$S(B)$ consists of two parts. c, an arbitrary constant, controls the relative emphasis of the two parts. Once an appropriate c is chosen $S(B)$ is minimized. The smaller the $S(B)$, the simpler the results. The most popular oblique rotations are quartimin method, $c = 0$, (Carroll, 1953), the covarimin method, $c = 1$, (Carroll, 1953; Kaiser, 1958, Harris and Kaiser (1964)) and biquartimin method with $c = 0.5$ (Carroll, 1957).

Using the popular quartimin method, the principal components from the same examples as those in Section 2.2.1 were rotated. As in Section 2.2.1, the oblique rotation is used on the first two components of RI data and Sparrow data and on the first three components of employment in Europe data.

For the RI data, the $S(B) = 0.05$ of rotated components is much less than the $S(B) = 1.17$ of unrotated components. So the oblique rotation makes the interpretation of components much easier. The important variables in the first rotated component are left-side Doppler RI and left-side CVI RI measurement (Table 2.4), i.e. measurements on the left side. The second rotated component mainly depends on right-side Doppler RI and right-side CVI RI measurement, i.e. measurements on the right side. The variances for rotated PCs are 3.47 and 3.35, for unrotated

Variables	Components 1	2
Right-side Doppler RI	0.07	-0.52
Left-side Doppler RI	0.51	-0.06
Right-side CVI RI	-0.08	-0.85
Left-side CVI RI	0.86	0.08
Angle	45°	42°
Variance	3.47	3.35
Cumulative variance	42%	82%

Table 2.4: Oblique rotated PCs for the RI data

are 4.81 and 2.15 respectively, the total variance for the rotated is slightly less than that of unrotated PCs (6.82 vs 7.02), and hence also slightly less than the total variance by orthogonal rotation. The first angle increases (45° vs 44°) and the second angle decreases (42° vs 44°) comparing with those by orthogonal rotation. For this example, it is not worth doing oblique rotation rather than orthogonal rotation. In oblique rotation, the two angles are different. This is possible because the axes are not required to remain perpendicular.

For the sparrow data, $S(B) = 0.08$ of rotated components is much less than unrotated PCs $S(B) = 2.68$. So by oblique rotation, the rotated components are very simple. The oblique rotated PCs are given in Table 2.5. Again the result is similar with that of orthogonal method. The variance explained by the first component (2.96) is less than that (3.62) of unrotated PCs, the variance explained by the second component (1.06) is more than that (0.53) of unrotated PCs. As the oblique results for RI data, the total variance of the rotated PCs is slightly less than that of unrotated PCs and orthogonal rotated PCs. The unimportant variable of component 1 is the length of keel of sternum, but the most important variable of component 2 is the length of keel of sternum. As in oblique rotation for RI data, the angles between the rotated and unrotated components are different. For the first component, the angle for the oblique rotation is more than that of orthogonal rotation (30° vs 28°), and for the second component, the angle

Variables	Components 1	2
Right-side Doppler RI	0.07	-0.52
Left-side Doppler RI	0.51	-0.06
Right-side CVI RI	-0.08	-0.85
Left-side CVI RI	0.86	0.08
Angle	45°	42°
Variance	3.47	3.35
Cumulative variance	42%	82%

Table 2.4: Oblique rotated PCs for the RI data

are 4.81 and 2.15 respectively, the total variance for the rotated is slightly less than that of unrotated PCs (6.82 vs 7.02), and hence also slightly less than the total variance by orthogonal rotation. The first angle increases (45° vs 44°) and the second angle decreases (42° vs 44°) comparing with those by orthogonal rotation. For this example, it is not worth doing oblique rotation rather than orthogonal rotation. In oblique rotation, the two angles are different. This is possible because the axes are not required to remain perpendicular.

For the sparrow data, $S(B) = 0.08$ of rotated components is much less than unrotated PCs $S(B) = 2.68$. So by oblique rotation, the rotated components are very simple. The oblique rotated PCs are given in Table 2.5. Again the result is similar with that of orthogonal method. The variance explained by the first component (2.96) is less than that (3.62) of unrotated PCs, the variance explained by the second component (1.06) is more than that (0.53) of unrotated PCs. As the oblique results for RI data, the total variance of the rotated PCs is slightly less than that of unrotated PCs and orthogonal rotated PCs. The unimportant variable of component 1 is the length of keel of sternum, but the most important variable of component 2 is the length of keel of sternum. As in oblique rotation for RI data, the angles between the rotated and unrotated components are different. For the first component, the angle for the oblique rotation is more than that of orthogonal rotation (30° vs 28°), and for the second component, the angle

Variables	Components 1	2
Total length	0.38	0.23
Alar extent	0.55	-0.08
Length of beak and head	0.55	-0.11
Length of humerus	0.50	0.03
Length of keel of sternum	-0.05	0.96
Angle	30°	23°
Variance	2.96	1.06
Cumulative variance	59%	80%

Table 2.5: Oblique rotated PCs for the Sparrow data

for the oblique rotation is less than that of orthogonal rotation (23° vs 28°).

The results from the oblique rotation for the employment in Europe data is given in Table 2.6. $S(B) = 1.52$ of rotated components is less than $S(B) = 2.51$ of unrotated PCs. So the rotation has improved the interpretation of PCs. The variances for the first two components are less than those of unrotated PCs (2.86 vs 3.49, 2.13 vs 1.90), and the variance for the third component is more than that of unrotated PCs (1.84 vs 1.10). So again, as with the oblique rotation results for RI data and sparrow data, the total variance of the rotated PCs is slightly less than that of unrotated PCs and orthogonal rotated PCs (6.60 vs 6.72). Each component obtained by oblique rotation is different from those obtained from orthogonal rotation. For the oblique rotated PCs, there are more variables of borderline importance. The first component is the difference between the agriculture, finance and manufacturing, construction, service industries, social and personal services, transport and communications. The second component represents the difference between mining, manufacturing, power supplies and social and personal services. The third component denotes the difference between agriculture, mining, transport and communications and power supplies, service industries, finance. The components obtained by oblique rotations do not make obvious improvement in interpretation of the PCs. Thus, as in

Variables	Components 1	2	3
1-Agriculture	-0.42	-0.16	-0.25
2-Mining	-0.10	0.57	-0.28
3-Manufacturing	0.22	0.46	-0.02
4-Power supplies	-0.07	0.61	0.29
5-Construction	0.36	0.05	-0.03
6-Service industries	0.26	-0.07	0.44
7-Finance	-0.24	0.04	0.72
8-Social and personal services	0.49	-0.22	0.07
9-Transport and communications	0.51	0.04	-0.26
Angle	31°	43°	48°
Variance	2.86	1.90	1.84
Cumulative variance	32%	53%	73%

Table 2.6: Oblique rotated PCs for the employment in Europe data

Variables	Components 1	2	3
Right-side Doppler RI	-0.22	-0.21	-0.72
Left-side Doppler RI	0.22	0.22	-0.69
Right-side CVI RI	0.05	-0.95	0.00
Left-side CVI RI	0.95	-0.05	-0.00
Angle	54°	44°	37°
Variance	2.81	2.69	2.24
Cumulative variance	34%	66%	94%

Table 2.7: Orthogonal rotated PCs for the RI data

Section 2.2.1, the oblique rotation is not successful for employment in Europe data.

Other popular types of oblique rotations applied to PCs for all three data obtain similar results. So changing the type of oblique rotation does not enhance the interpretation of PCs.

2.2.3 Some comments about rotation

If rotation is used, various questions should be addressed. Should the rotation be orthogonal or oblique? Which simplicity criteria should be chosen? How many components should be rotated? The idea of using rotation of the loadings to simplify their interpretation is widely used but controversial (Richman, 1986, 1987; Jolliffe, 1987, 1995; Mestas-Nunez 2000). A drawback of rotation is that rotated components are not invariant to a change in the number of components (Jackson, 1991). A change from m to $m+1$ may change the nature of all the rotated PCs, because the rotation is now in $m+1$ space rather than in m space.

Take the varimax rotation results for the RI data as an example. When $m = 2$, the rotation results and interpretation are as listed in the Section 2.1.1. When $m = 3$, the rotation results are as given in Table 2.7. For this data, when $m = 3$, $S(B)$ is 1.39 for rotated components. This is much more than that ($S(B) = 0.14$) for original PCs. So that the rotated components

are more simple than the unrotated components. Thus when $m = 3$ the interpretation of the components is very easy. When $m = 3$, the first rotated component largely represents the CVI-L measurement with a little bit of the contrast between measurements on the Doppler machine. The second rotated component represents the CVI-R measurement (mainly on the right side of CVI RI measurement) and in a minor way the contrast between measurements on the Doppler machine. The third component denotes the measures of using Doppler machine. The angles between the original and rotated components are different, this is possible for orthogonal rotation because $m = 3$, the size of the angles are too big.

Next, compare this results with the results in Table 2.1. The simplicity criterion for the first two components of RI data in Table 2.7 is 1.14. This is much bigger than that ($S(B) = 0.69$) when $m = 2$ in Table 2.1. So the first two rotated components in Table 2.7 are more simple than the rotated components in Table 2.1. The first two components using $m = 3$ and $m = 2$ account for 66% and 84% of the variance respectively. So for more simple components, there is 18% loss in variance for the first two components. Furthermore the angles between the rotated and original components when $m = 2$ in Table 2.1 are not more than those of the orthogonal results when $m = 3$ (44° vs 54°, 44° vs 44°). So the first two rotated principal components and the nature of all the rotated components have changed when using $m = 3$ instead of $m = 2$.

Orthogonal rotation generally is used since orthogonal matrices are easy to compute with, hence orthogonal rotation costs less than oblique rotation. But oblique rotation is used for several reasons. It is unlikely that influences in nature are uncorrelated. Even if they are uncorrelated in the population, they need not be so in the sample. Thus, oblique rotations have often been used to get easier interpretation components in practice. So overall rotation methods can be widely used to enhance the interpretation of principal components because of the easy use and readily available software (for example, S-Plus), but before using them, you must decide how many components to take.

2.3 Truncation of loadings

There are a lot of procedures for discarding or selecting variables in PCA, factor analysis and regression. When these procedures are applied to PCA, they are equivalent to truncation of the loadings of PCs, i.e. setting small loadings of PCs to zero so that the interpretation of components becomes easier. These procedures are divided into two types. One type is based on the loadings of PCs, the loadings whose absolute values are sufficiently small are set to zero. The other type is to drop variables that do not seem to be making much contribution across the board, this type of method will not talked about in this thesis.

The first type of approach is popularly used in practice. This approach treats all the loadings smaller than some threshold absolute value as zero. Take the analysis of Jeffers' Pitprop data (originally analysed by Jeffers (1967) and discussed in Section 1.3.4) as an example. Recall that for the pitprop data $m = 4$ (Section 1.3.4). Checking the first four PCs obtained for this data set in Table 1.9, the loadings less than 0.2 are relatively small compared to the largest loading in each component. So 0.2 is chosen as the threshold for all the retained components in this example.

The new vectors are not orthogonal after truncation (Table 2.8). There is loss in variance explained by each component compared to first PCs in Table 1.9, the cumulative variance accounted for becomes 70% compared to 74% of the first four PCs. In caculating variances it is not the loadings in Tables 2.8 and 2.9 that are used, but normalised versions of them. The angles between the truncated PCs and original PCs are 16, 21, 13 and 19 degrees respectively. The first component is a weighted average of the top diameter of the prop, the length of the prop, the number of annual rings at the top and the base of the prop, the maximum bow, the distance of the point of maximum bow from the top of the prop and the number of knot whorls. So it is a measure of the size of the prop. The second component appears to be a difference between the top diameter of the prop, the moisture content of the prop, the specific gravity of the timber, the length of clear prop from the top of the prop, the average number of knots per whorls, the average diameter of the knots and the number of knot whorls. The third component represents a difference between the top diameter of the prop, the length of the prop, the maximum bow,

Variables	Components			
	1	2	3	4
1-Topdiam	-0.40	0.22	-0.21	0
2-Length	-0.41	0	-0.24	0
3-Moist	0	0.54	0	0
4-Testsg	0	0.46	0.35	0
5-Ovensg	0	0	0.48	0
6-Ringtop	-0.28	0	0.48	0
7-Ringbot	-0.40	0	0.25	0
8-Bowmax	-0.29	0	-0.24	0.29
9-Bowdist	-0.36	0	-0.21	0
10-Whorls	-0.38	-0.25	0	-0.21
11-Clear	0	0.21	0	0.80
12-Knots	0	0.34	0	-0.30
13-Diaknot	0	0.31	-0.33	-0.30
Angle	16°	21°	13°	19°
Variance	3.99	2.15	1.84	1.10
Cumulative variance	31%	48%	62%	70%

Table 2.8: Truncation of loadings (threshold 0.2) for the pitprop data

Variables	Components			
	1	2	3	4
1-Topdiam	-0.40	0.22	-0.21	0
2-Length	-0.41	0	-0.24	0
3-Moist	0	0.54	0	0
4-Testsg	0	0.46	0.35	0
5-Ovensg	0	0	0.48	0
6-Ringtop	-0.28	0	0.48	0
7-Ringbot	-0.40	0	0.25	0
8-Bowmax	-0.29	0	-0.24	0.29
9-Bowdist	-0.36	0	-0.21	0
10-Whorls	-0.38	-0.25	0	-0.21
11-Clear	0	0.21	0	0.80
12-Knots	0	0.34	0	-0.30
13-Diaknot	0	0.31	-0.33	-0.30
Angle	16°	21°	13°	19°
Variance	3.99	2.15	1.84	1.10
Cumulative variance	31%	48%	62%	70%

Table 2.8: Truncation of loadings (threshold 0.2) for the pitprop data

the distance of the point of maximum bow from the top of the prop, the average diameter of the knots per whorls and the specific gravity of the timber, the oven-dry specific gravity of the timber, the number of annual rings at the top and at the base of the prop. The fourth component represents a difference between the maximum bow and the length of clear prop from the top of the prop on one side, and the number of knot whorls, the average number of knots per whorls, the average diameter of the knots on other side. The variable Clear seems very important.

The first four components don't have a nice interpretation except for component 1. So 0.2 is not a good choice of threshold to improve the interpretation of pitprop data. The interpretation of components is simple but not simple enough.

Using a threshold of 0.4 gives the results in Table 2.9. The interpretation is straightforward. The first component is an average of the top diameter of the prop, the length of the prop and the number of annual rings at the base of the prop. The second component is a weighted average of the moisture content of the prop and the specific gravity of the timber. The third component represent the average of the oven-dry specific gravity of the timber and the number of annual rings at the top of the prop. The fourth component is the length of clear prop from the top of the prop. The larger the threshold, the easier the interpretation generally but the larger the angle (45°, 45°, 47°, 37° vs 16°, 21°, 13°, 19°) between the truncated component and corresponding principal component. This is expected because large threshold leads to more zero loadings of truncated components so that the interpretation of truncated components is easier. Also more zero truncated loadings means that more loadings differ from the loadings of the original PCs, thus the angle between original PC and corresponding truncated PCs becomes bigger. Furthermore, the cumulative variance explained by the first four components is 50%. This is less than the 64% explained by the first four components when threshold is 0.2 and 74% explained by the original PCs. The biggest loss in variance appears in component 1, from 33% to 18%, there is 15% loss in variance. So 0.4 is too big, in particular for the first component because it's almost as big as the biggest loading of the first component. So, I don't think 0.4 is a good choice for this data because too much cumulative variance explained is lost. Jeffers (1967) investigated the Pitprop data. He ignored the loadings whose absolute values less than the 70% of the absolute value of maximum loadings of each component. So each component has

Variables	Components			
	1	2	3	4
1-Topdiam	-0.40	0	0	0
2-Length	-0.41	0	0	0
3-Moist	0	0.54	0	0
4-Testsg	0	0.46	0	0
5-Ovensg	0	0	0.48	0
6-Ringtop	0	0	0.48	0
7-Ringbot	-0.40	0	0	0
8-Bowmax	0	0	0	0
9-Bowdist	0	0	0	0
10-Whorls	0	0	0	0
11-Clear	0	0	0	0.80
12-Knots	0	0	0	0
13-Diaknot	0	0	0	0
Angle	45°	45°	47°	37°
Variance	2.30	1.87	1.36	1.00
Cumulative variance	18%	32%	42%	50%

Table 2.9: Truncation of loadings (threshold 0.4) for the pitprop data

a different threshold. For example, the first component, the maximum loading is 0.41, so for this component the threshold is about 0.29. For component 2, the threshold is about 0.38. For component 3, the threshold is about 0.33. For component 4, the threshold is about 0.56. Every threshold chosen by Jeffers is bigger than 0.2. That's why his interpretation for the truncated components is much easier than the truncated components in Table 2.8. For components 1 and 3, the interpretation with threshold 0.4 is easier than that of Jeffers. For components 2 and 4, the interpretation is the same (for component 2, no loading is between 0.38 (Jeffers' threshold) and 0.4, for component 4, not loading between 0.4 and 0.56 (Jeffers' threshold). So choosing a suitable threshold is very important. Too small a threshold may be not useful to improve interpretation of PCs, too big a threshold may lead to a big difference from the original principal component.

The truncating approach used in the above example is to ignore all the small loadings of PCs. Chipman and Gu (2003) give another procedure for setting the smaller loadings to zero, more details of this approach will be given in Section 2.4. This type of truncation of loadings to some extent overcomes the drawbacks of the previous procedure. However this approach implicitly assumes that the size (absolute value) of the loadings for a PC are a reliable guide to interpreting that PC. However variables with the smallest loadings are not necessarily the least important in representing a component (Cadima and Jolliffe, 1995). According to Cadima and Jolliffe's viewpoint, if the covariance matrix PCs are considered, it is not just the loadings of the linear combination but also the variance and the correlations between the variables which determine the results.

Truncating PCs can be considered as choosing a subset of variables (those not truncated) with which to approximate a PC. So any approach for selecting a subset of variables in PCA can be used as the methods truncating the loadings of PCs. Recall that the same interpretation issue as PCA arises in multiple regression, where the response is a linear combination of the predictor. The linear models are obtained via variable selection. So the approaches applied in multiple regression can be applied to PCA. In particular, the approaches to drop variables that do not seem to be making much contribution across the board can be applied to PCA, this type of approach both considers the variance and the correlation between the variables such as

the application in the regression area (Hocking 1976, 1983), McCabe (1984) (discussed later), Jolliffe (1972, 1973). Beale, Kendall and Mann (1967) introduce an approach to discarding variables called interdependence analysis. They suggest maximizing the minimum multiple correlation between the selected variables and discarded variables. So these approaches overcome the drawback of above methods. I will not give more details of these methods because these methods are not immediately used to simplify PCs.

2.4 Restriction of loadings to a discrete set of values

Some approaches find linear combinations of p variables that maximize variance, as in PCA, but limit the number of available values for the loadings to a small set to aid interpretation. The idea dates back to Hausmann (1982) who allowed the loadings only to take the value -1, 0, 1. Take the first component of RI data as an example. The first principal component is $(0.42, 0.43, 0.55, 0.58)$, so there are 40 possible loadings for the first component ($C_4^1 + 2 \times C_4^2 + 2^2 \times C_4^3 + 2^3 \times C_4^4 = 40$). Vector $(0, 0, 0, 1)$ is the vector with maximal variance (2.66) among $C_4^1 = 4$ vectors with 1 nonzero loadings. Vector $(0, 0, 1, 1)$ is the vector with maximal variance (3.36) among $2 \times C_4^2 = 12$ vectors with 2 nonzero loadings. Vector $(1, 0, 1, 1)$ is the vector with maximal variance (4.05) among $2^2 \times C_4^3 = 16$ vectors with 3 nonzero loadings. Vector $(1, 1, 1, 1)$ would be the directions with maximal variance (4.73) among $2^3 \times C_4^4 = 8$ with 4 nonzero loadings. Comparing all of the corresponding variances, the one with the highest variance is $(1, 1, 1, 1)$. So the modified PC obtained by Hausmann's method is $(1, 1, 1, 1)$. As these can be derived from the variance covariance matrix without knowing the PCs, it's a one stage approach. For Hausman's method, as with the trucating loadings, in calculating variances it is not the loadings that are used, but normalised versions of them.

Simple components defined by Vines (2000) require the loadings of simple components to be integers, this method will be discussed in the next chapter.

Chipman and Gu (2003) first obtain ordinary PCs, and then constrained loadings. So this is a two stage approach. In their method, to enhance interpretation of PCs, they provided three classes of constraint, a homogeneity constraint, a contrast constraint and a sparsity constraints.

First, consider the homogeneity constraint. As in Section 1.2, suppose the loadings of the k^{th} component $W_k = x\gamma_k$ $(k = 1, \ldots, p)$ take very few distinct values, say 0, $-c$ and c, $c > 0$, such that $\gamma_k^T \gamma_k = 1$, where $\gamma_{k,j}$ is the j^{th} loading of γ_k. If a PC is an average of some of the variables, the homogeneity constraint is used. In this case, the best γ_k is defined as the component with the smallest angle with k^{th} PC. The search algorithm is straightforward. Among all γ_k with i nonzero loadings, $i = 1, 2, \ldots, p$, identify the i loadings of k^{th} PC with largest absolute values. Set the corresponding loadings of γ_k to $1/\sqrt{i}$ or $-1/\sqrt{i}$ according to the signs of the i^{th} loadings of k^{th} PC. All other elements of γ_k are zero so that $\gamma_k^T \gamma_k = 1$. The best γ_k is obtained by repeating this procedure for $i = 1, \ldots, p$. So there only p possible vectors are checked.

For example, the first principal component of RI data is $(0.42, 0.43, 0.55, 0.58)$. Thus the vectors $(0, 0, 0, 1)$, $(0, 0, 1, 1)/\sqrt{2}$, $(0, 1, 1, 1)/\sqrt{3}$ and $(1, 1, 1, 1)/2$ would be the vectors that would be checked. The angles between these vectors and the first principal component are 55°, 37°, 26° and 8° respectively. In this case, $(1, 1, 1, 1)/2$ is the closest to this PC.

The contrast constraint generally is used when a PC is a difference of the average of one set of variables and the average of another set of variables. The contrast constraint limits loadings to 0, $-c_1$ and c_2 for positive c_1 and c_2 such that $\sum_{j=1}^{p} \gamma_{kj} = 0$ and $\gamma_k^T \gamma_k = 1$. The algorithm for identifying contrast constraint is very similar for homogeneous constraint. One minor difference is that the vector must contain at least one element of each sign because the sum of all the elements in a vector must be zero. Consider the second component of RI data $(-0.32, 0.30, -0.65, 0.63)$. The vectors $(0, 0, -1, 1)/\sqrt{2}$, $(-1, 0, -1, 2)/\sqrt{6}$, $(0, 1, -2, 1)/\sqrt{6}$ and $(-1, 1, -1, 1)/2$ would be the best γ with $i = 2, 3, 4$ nonzero loadings. The angles between these vectors and the second principal component are 25°, 24°, 24° and 18° respectively. In this case, $(-1, 1, -1, 1)/2$ is the closest to the second PC.

A component with many zero loadings is called a sparse component. The sparsity constraint attempts to set as many loadings to zero as possible. So the sparse constraint is used if the components have several loadings near zero, while others take a wide range of values. To find sparse components which approximate principal components well, the angle between the sparse component and the corresponding principal component is minimized. The angle is minimized when no loadings are zero, so a penalty term is added. Thus $C_1 = \theta/(\pi/2) + \eta i/p$ is minimized,

First, consider the homogeneity constraint. As in Section 1.2, suppose the loadings of the k^{th} component $W_k = x\gamma_k$ $(k = 1, \ldots, p)$ take very few distinct values, say 0, $-c$ and c, $c > 0$, such that $\gamma_k^T \gamma_k = 1$, where $\gamma_{k,j}$ is the j^{th} loading of γ_k. If a PC is an average of some of the variables, the homogeneity constraint is used. In this case, the best γ_k is defined as the component with the smallest angle with k^{th} PC. The search algorithm is straightforward. Among all γ_k with i nonzero loadings, $i = 1, 2, \ldots, p$, identify the i loadings of k^{th} PC with largest absolute values. Set the corresponding loadings of γ_k to $1/\sqrt{i}$ or $-1/\sqrt{i}$ according to the signs of the i^{th} loadings of k^{th} PC. All other elements of γ_k are zero so that $\gamma_k^T \gamma_k = 1$. The best γ_k is obtained by repeating this procedure for $i = 1, \ldots, p$. So there only p possible vectors are checked.

For example, the first principal component of RI data is $(0.42, 0.43, 0.55, 0.58)$. Thus the vectors $(0, 0, 0, 1)$, $(0, 0, 1, 1)/\sqrt{2}$, $(0, 1, 1, 1)/\sqrt{3}$ and $(1, 1, 1, 1)/2$ would be the vectors that would be checked. The angles between these vectors and the first principal component are 55°, 37°, 26° and 8° respectively. In this case, $(1, 1, 1, 1)/2$ is the closest to this PC.

The contrast constraint generally is used when a PC is a difference of the average of one set of variables and the average of another set of variables. The contrast constraint limits loadings to 0, $-c_1$ and c_2 for positive c_1 and c_2 such that $\sum_{j=1}^{p} \gamma_{kj} = 0$ and $\gamma_k^T \gamma_k = 1$. The algorithm for identifying contrast constraint is very similar for homogeneous constraint. One minor difference is that the vector must contain at least one element of each sign because the sum of all the elements in a vector must be zero. Consider the second component of RI data $(-0.32, 0.30, -0.65, 0.63)$. The vectors $(0, 0, -1, 1)/\sqrt{2}$, $(-1, 0, -1, 2)/\sqrt{6}$, $(0, 1, -2, 1)/\sqrt{6}$ and $(-1, 1, -1, 1)/2$ would be the best γ with $i = 2, 3, 4$ nonzero loadings. The angles between these vectors and the second principal component are 25°, 24°, 24° and 18° respectively. In this case, $(-1, 1, -1, 1)/2$ is the closest to the second PC.

A component with many zero loadings is called a sparse component. The sparsity constraint attempts to set as many loadings to zero as possible. So the sparse constraint is used if the components have several loadings near zero, while others take a wide range of values. To find sparse components which approximate principal components well, the angle between the sparse component and the corresponding principal component is minimized. The angle is minimized when no loadings are zero, so a penalty term is added. Thus $C_1 = \theta/(\pi/2) + \eta i/p$ is minimized,

Variables	Components 1	2
Total length	$1/\sqrt{5}$	0
Alar extent	$1/\sqrt{5}$	$-1/2\sqrt{3}$
Length of beak and head	$1/\sqrt{5}$	$-1/2\sqrt{3}$
Length of humerus	$1/\sqrt{5}$	$-1/2\sqrt{3}$
Length of keel of sternum	$1/\sqrt{5}$	$3/\sqrt{12}$
Angle	4°	5°
Variance	3.61	0.54
Cumulative variance	72%	83%

Table 2.10: Chipman and Gu's method applied to PCs for the sparrow data

where i is the number of nonzero loadings in γ_k, θ is the angle between sparse component $x\gamma_k$ and the k^{th} PC and η is a tuning parameter. As η increases, the component minimizing C_1 becomes more sparse. Other criteria are possible. For example, one could maximize $C_2 = (p-i)(cos\theta)^\eta$. The $(p-i)$ term is large when most loadings are zero and the $\cos\theta$ is large if θ is small. The C_1 criterion could select the original principal component in some case (maybe C_1 is minimized when the selected component is principal component). Such a selection is impossible for C_2 because in this case $C_2 = 0$, it can't be the best vector. Otherwise these two criteria behave similarly. So the search for the sparse components is similar to the homogeneous case. For example, if a component is $(0.05, -0.03, 0.5, 0.86)$, when $\eta = 0.8$, $(0,0,0,1)$, $(0,0,1,1)/\sqrt{2}$, $(1,0,1,1)/\sqrt{3}$ and $(1,-1,1,1)/2$ would be the best γ with $i = 1,2,3,4$ nonzero loadings. The corresponding C_1 for these vectors are 0.54, 0.58, 0.99 and 1.29 respectively, the minimal C_1 is obtained by the first vector, so the best vector for PC is $(0,0,0,1)$. The corresponding C_2 for these vectors are 2.66, 1.94, 0.85 and 0, the maximal C_2 is obtained by the first vector. Again the best vector is $(0,0,0,1)$. So these two criteria behave similarly for this example.

The homogeneity constraint and contrast constraint were applied to the first two PCs of sparrow data in Table 1.4. Checking the principal components in Table 1.4, the first PC is an average

of the variables, so the homogeneity constraint is used. The second PC is a difference between length of keel of sternum and alar extent, length of beak and head, length of humerus, so the contrast constraint is used. For the first component, the vectors $(0,0,0,1,0)$, $(0,1,0,1,0)/\sqrt{2}$, $(1,1,0,1,0)/\sqrt{3}$ and $(1,1,1,1,0)/2$ and $(1,1,1,1,1)/\sqrt{5}$ would be the best γ with $i = 1,2,3,4,5$ nonzero loadings. The angles between these vectors and the first principal component are 62°, 49°, 37°, 24° and 4° respectively. So $(1,1,1,1,1)/\sqrt{5}$ is the closest to the first PC. For the second component, the vectors $(0,0,-1,0,1)/\sqrt{2}$, $(0,-1,-1,0,2)/\sqrt{6}$, $(0,-1,-1,-1,3)/\sqrt{12}$ and $(1.5,-1,-1,-1,1.5)/\sqrt{7.5}$ would be the best vectors with $i = 2,3,4,5$ nonzero loadings. The angles between these vectors and the second principal component are 31°, 13°, 5° and 37° respectively. So $(0,-1,-1,-1,3)/\sqrt{12}$ is the closest to the second PC. The results are given in Table 2.10. The interpretation of the results in Table 2.10 is straightforward. The first component is the average of all the variables. The second component represents the contrast between the length of keel of sternum and the average of alar extent, length of beak and head, length of humerus. The first variance slightly decreases (3.61 vs 3.62) and the second variance slightly increases (0.53 vs 0.54) compared to PCA in Table 1.4. So the cumulative variance explained by the first two components is the same as that by the first two PCs (83%). Next, comparing the results in Table 2.10 with the orthogonal and oblique rotation results for this data in Tables 2.2 and 2.5, the cumulative variance for these three methods are all the same (83%). But the orthogonal rotation (2.93 vs 3.62) and the oblique rotation (2.96 vs 3.62) lost too much variance for the first component. Furthermore the angle between the first component and the first PC is 4 degrees but for orthogonal rotated PCs the angle is 29 degrees and for oblique rotated PCs the angle is 30 degrees (Table 2.5). The angle between the second component and the second PC is 4 degrees but for orthogonal rotated PCs the angle is 28 degrees and for oblique rotated PCs the angle is 23 degrees. So the results here are easier to interpret and the angles are much smaller. Thus for this example, the results in Table 2.10 are much better than the rotated PCs in Tables 2.2 and 2.5.

The components obtained by Chipman and Gu (2003) (Table 2.10) are very good for this data. However, sometimes it is difficult to decide which constraint to use for a given component. For example, vector $(0.41,-0.03,-0.42,0.81)$, I am not sure whether to use the

homogeneity constraint or contrast constraint. For the second component of sparrow data, $(0.05, -0.30, -0.33, -0.18, 0.88)$, it is also not very clear which constraint to use. When the homogeneity constraint is used in this component, the best component corresponds to $(0, 0, 0, 0, 1)$. This is much different from the vector in Table 2.10 when contrast constraint is used. So the best results from different constraint must be considered. It's not an easy process.

2.5 LASSO-based approaches

Tibshirani (1996) proposed the "Least Absolute Shrinkage and Selection Operator", LASSO. The LASSO approach was developed in multiple regression analysis to deal with multicollinearity, or simply a large number of variables included in the regression equation. The LASSO is a variable selection technique. Suppose the standard multiple regression is $Y = \sum_{j=1}^{p} X_j\beta_j$, $i = 1, 2, \ldots, n$, where $Y = (y_1, y_2, \ldots, y_n)$ is the response, $X_j = (x_{1j}, \ldots, x_{nj})^T$, $j = 1, 2, \ldots, p$ are corresponding values of p predictor variables, and $\beta_1, \beta_2, \ldots, \beta_p$ are parameters in the regression equation. In least squares regression, these parameters are estimated by minimizing the residual sum of squares, $\sum_{i=1}^{n}(Y - \sum_{j=1}^{p} X_j\beta_j)^2$. The aim of the LASSO approach is to make some of the regression coefficients become exactly zero. To achieve this an extra constraint is added to the coefficients, $\sum_{j=1}^{p} |\beta_j| \leq t$ for some 'tuning parameter' t. Suitable choices of t forces the coefficients in the regression equation to zero. An equivalent way of deriving LASSO estimates is to minimize $\sum_{i=1}^{n}(Y - \sum_{j=1}^{p} X_j\beta_j)^2 + \lambda\sum_{j=1}^{p} |\beta_j|$, where λ is a non-negative value. For any given value of t in the first LASSO formulation there is a value of λ in the second formulation that gives the same results.

Jolliffe and Uddin (2002) applied LASSO to PCA (SCoTLASS). As with PCA, their idea is to maximize the variance $\gamma_k^T S\gamma_k$, subject to $\gamma_k^T\gamma_k = 1$, for $k \geq 2$ and $\gamma_h^T\gamma_k = 0$, $h < k$. The constraint $\sum_{j=1}^{p} |\gamma_{kj}| \leq t$, where γ_{kj} is the j^{th} loading in the k^{th} vector $X\gamma_k$, $k = 1, \ldots, p$ and t is a tuning parameter that lies between 1 and $\sqrt{p}$, is imposed. This constraint drives some loadings α_{kj} to zero as t decreases from $\sqrt{p}$. Unlike rotation, this technique usually produces some exactly zero loadings in the components. As with the LASSO approach, SCoTLASS is equivalent to minimizing $\gamma_k^T S\gamma_k + \lambda\sum_{j=1}^{p} |\gamma_{kj}|$ subject to $\gamma_k^T\gamma_k = 1$, and for $k \geq 2$ $\gamma_h^T\gamma_k = 0$,

$h < k$. where λ is a non-negative value. So SCoTLASS is a one stage approach.

Zou and Hastie (2003) noticed that the LASSO has limitations. For example, the number of selected variables by the LASSO is limited by the number of observations. Suppose there are thousands of predictors (p) but less than one hundred in the sample (n), the LASSO can only select at most n predictors. So Zou and Hastie (2003) introduce the elastic net to generalizes the LASSO to overcome this drawback, while having similar optimal properties. For any non-negative λ and λ_1, the elastic net estimates are obtained by minimizing

$(1+\lambda)\sum_{i=1}^{n}(Y-\sum_{j=1}^{p}X_j\beta_j)^2+\lambda\sum_{j=1}^{p}|\beta_j|^2+\lambda_1\sum_{j=1}^{p}|\beta_j|$

LASSO is the special case when $\lambda = 0$. When $p > n$, λ is positive, then the elastic net can include all variables in the fitted model, so the limitation of the LASSO is overcome.

Zou, Hastie and Tibshirani (2004) introduced a different approach called sparse principal component analysis (SPCA) to modify PCA. This approach can directly make use of the LASSO to produce modified principal components with sparse loadings, i.e. use LASSO make some of the loadings of PCs become zero. So the components obtained by SPCA are called components with sparse loadings. In this description of SPCA, direct sparse approximations are introduced first. Each PC is a linear combination of the p variables, thus its loadings can be recovered by regressing the PC on the p variables. As in Section 1.2, $Z_i = U_iL_i$ is the i^{th} principal component, where U is a $n \times r$ matrix whose k^{th} column is k^{th} PCs with unit length, r is the rank of X. The variance of the k^{th} PC is L_{kk}^2, L is an $r \times r$ diagonal matrix. The new elastic net $\hat{\beta}$ estimates are obtained by minimizing $|Z_i - X\beta|^2 + \lambda|\beta|^2 + \lambda_1|\beta|_1$, where $|\beta|^2 = \sum_{j=1}^{p}\beta_j^2$, $\beta_1 = \sum_{j=1}^{p}|\beta_j|$. Then $\hat{A}_i = \hat{\beta}/|\hat{\beta}|$ an approximation to A_i, and $X\hat{A}_i$ is the i^{th} approximated PC. So the above approach is a two stage approach. First perform PCA, then use above formulation to find suitable sparse approximations. The new elastic net estimates in above formulation differs from the elastic net by a scaling factor $(1 + \lambda)$. Since the normalized fitted coeffients are used, the scaling factor does not affect $\hat{A}_i$.

Next, Zou, Hastie and Tibshirani (2004) presented a "self-contained" regression-type criterion to derive PCs. When X_i denotes the i^{th} row vector of matrix X, the previous formulation becomes

$\sum_{i=1}^{n}(X_i - \alpha\beta^T X_i)^2 + \lambda\sum_{j=1}^{r}|\beta_j|^2 + \sum_{j=1}^{r}\lambda_{1,j}|\beta_j|_1$

subject to $\alpha^T\alpha = I_r$ and $\hat{A}_i = \hat{\beta}/|\hat{\beta}|$ for $i = 1, \ldots, r$.

where α and β be $p \times r$ matrices, the same λ is used for all r PCs (as in Section 1.2, r is the rank of matrix X), different $\lambda_{1,j}s$ are allowed for penalizing the loadings of different principal components. If $p > n$, a positive λ is required to get PCA. When $\lambda_{1,j} = 0$, the LASSO penalty vanishes.

Finally, SCoTLASS and SPCA approaches are applied to the pitprop data. The results (Jolliffe and Uddin (2002)) obtained by SCoTLASS are presented for only one value of the tuning parameter $t = 1.75$, and it is not easy to choose the t. As before (Section 1.3.4), only the first four components are retained (Table 2.11).

The results for Jeffers' pitprop data using SPCA (Zou, Hastie and Tibshirani (2004)) are given in Table 2.12. As before (Section 1.3.4), only the first four components are retained. The PCs by SCoTLASS (Table 2.11) and the PCs by SPCA (Table 2.12) account for 67% and 65% variance respectively. These are less than the cumulative variance for the first four components by the original PCs in Table 1.9 (74%). The angles produced by SCoTLASS and SPCA are very similar except angle 4 (65° vs 37°). The angles of the truncated results using a threshold of 0.2 (Table 2.8) are smaller than those given by SPCA and SCoTLASS, and the angles of the truncated results using a threshold of 0.4 (Table 2.9) are larger than those given by SPCA and SCoTLASS. So the truncated components in Table 2.8 (threshold 0.2) are more close to the original PCs than those given by SPCA and SCoTLASS. For first component, the interpretation is similar for the PC by SCoTLASS and PC by SPCA, but the non-zero loadings by SPCA for Ovensg confuses things. The interpretations for components 2, 3 and 4 obtained by SCoTLASS and by SPCA are relatively easier to interpret than the PCs (truncated using a threshold of 0.2) in Table 2.8 (explained 64% of the variance) and original PCs in Table 1.9 because the PCs obtained by SPCA and SCoTLASS have fewer nonzero loadings, especially the PCs obtained by SPCA. The truncated results in Table 2.9 (which use a threshold of 0.4) have more zero loadings but it explains too little variance (50%). So the results by SPCA are considered the best among the results in Tables 2.8, 2.9, 2.11 and 2.12. So only the interpretation of PCs obtained by SPCA (Table 2.12) are given. Once again, the first component is a measure of size of the prop. The second component is the average of the moisture content of the prop and the specific gravity

t=1.75 Variables	Components			
	1	2	3	4
1-Topdiam	0.55	0.05	-0.09	0.07
2-Length	0.57	0.00	-0.08	0.12
3-Moist	0.00	0.64	-0.19	-0.13
4-Testsg	0.00	0.64	0.00	-0.14
5-Ovensg	0.00	0.00	0.46	0.00
6-Ringtop	0.00	0.36	0.35	0.00
7-Ringbot	0.28	0.00	0.33	0.00
8-Bowmax	0.13	-0.01	0.00	-0.59
9-Bowdist	0.38	0.00	0.00	0.00
10-Whorls	0.38	-0.07	0.00	-0.07
11-Clear	0.00	0.00	0.00	0.00
12-Knots	0.00	0.21	0.00	0.77
13-Diaknot	0.00	0.00	-0.72	0.01
Number of nonzero loadings	6	7	7	8
Angle	28°	43°	44°	65°
Variance	27%	16%	15%	9%
Cumulative Variance	27%	43%	58%	67%

Table 2.11: SCoTLASS for the pitprop data

Variables	Components			
	1	2	3	4
1-Topdiam	-0.48	0.00	0.00	0
2-Length	-0.48	0.00	0.00	0
3-Moist	0.00	0.79	0.00	0
4-Testsg	0.00	0.62	0.00	0
5-Ovensg	0.18	0.00	0.64	0
6-Ringtop	0.00	0.00	0.59	0
7-Ringbot	-0.25	0.00	0.49	0
8-Bowmax	-0.34	-0.02	0.00	0
9-Bowdist	-0.42	0.00	0.00	0
10-Whorls	-0.40	0.00	0.00	0
11-Clear	0.00	0.00	0.00	-1
12-Knots	0.00	0.01	0.00	0
13-Diaknot	0.00	0.00	-0.02	0
Number of nonzero loadings	7	4	4	1
Angle	28°	44°	44°	37°
Variance	28%	14%	15%	8%
Cumulative Variance	28%	42%	57%	65%

Table 2.12: SPCA for the pitprop data

of the timber at the time of test. It seems to be a measure of the density of the timber at the time of test. The third component represents the average of the oven-dry specific gravity of the timber and the numbers of annual rings at the top and base of the prop. The fourth component is the length of clear prop from the top of the prop.

2.6 Combining variance maximization (improvement maximization in variance) and simplification

The rotation methods described in Section 2.2 do not get 'simple components" directly. As has already been noted, rotation methods have two stages. They first obtain the PCs and then rotate them. A lot of approaches which are used to enhance the interpretation of PCs combine variance maximization and simplification (Jolliffe, Uddin and Vines (2002)). So these are one stage approaches such as SCoTLASS and SCA described in the previous section.

Jolliffe and Uddin (2000) with their simplified component technique (SCoT) also combine the variance maximization and simplicity by introducing a penalty function, the penalty function (simplicity criteria) $Sim(\gamma_k)$ is introduced to make the linear combination simple. As in Section 1.2, suppose the k^{th} simplified component $W_k = X\gamma_k$, X is a $n \times p$ matrix, n is the number of observation and p is the number of variables. X_i is the i^{th} row of X and $X_i\gamma_k$ is the value of W_k for the i^{th} observation. The penalty function (simplicity) is defined as the varimax criterion, as defined in equation (2.1) when $c = 1$. If $V(\gamma_k) = Var(X\gamma_k)$, $Var(X\gamma_k)$ denotes the sample variance $\gamma_k^T S\gamma_k$ of $X\gamma_k$ then the SCoT successively maximizes $(1-\Phi)V(\gamma_k)+\Phi Sim(\gamma_k)$ subject to $\gamma_k^T\gamma_k = 1$ and, for $k \geq 2$, $\gamma_k^T\gamma_j = 0$, $j < k$. Here Φ is a tuning parameter, which needs to be chosen. The value $\Phi = 0$ corresponds to PCA. As Φ increases, the components of SCoT move away from the PCs but become more simple. When $\Phi = 1$, each component from SCoT is identical to one original variable.

SCoT does not usually agree with two stage rotated PCA. Rotated PCs are defined within the subspace obtained by the m retained PCs, so that the m-dimensional subspace has maximum variance. This method sacrifices some variance in the process of searching for simplicity, since

the introduction of the penalty function takes the calculation out of this subspace.

McCabe (1984) proposed the principal variables approach. This approach simply looks for the best q original variables, termed "principal variables", with respect to various criteria. Suppose we have obtained a data matrix Y. Assume Y is divided into an $n \times q$ submatrix $Y1$ and a $n \times (p-q)$ submatrix $Y2$ $(1 < q < p)$ and $Y = (Y1, Y2)$. Then one of the following four criteria, $C1$, ..., $C4$, are used to find a subset of the original variables called principal variables. $C1 = mindet(S_{22.1})$, $C2 = mintr(S_{22.1})$, $C3 = min||S_{22.1}||$,C4=max cancor(Y_1, Y_2), where $S_{22.1}$ is the conditional covariance (or correlation) matrix of the $Y2$, given the $Y1$, $tr(S_{22.1})$ represents the trace of the matrix $S_{22.1}$, $||S_{22.1}||$ denotes the Euclidean norm of matrix $S_{22.1}$ and equals the square root of the sum of squares of all the loadings in the matrix, $cancor(Y1, Y2)$ is the canonical correlation between the set of $Y1$ and $Y2$. This idea can be used in PCA to reduce the variables in each PC, the principal variables (Y_1) of the original variables are retained, while the others are ignored (their loadings are taken as 0), so the PCs are easier to interpret. The principal variables are chosen after doing the PCA, so this is a two stage approach.

It is desirable that approaches not only enhance the interpretation but also have an explicit definition of simplicity. The method provided by Rousson and Gasser (2004) termed SCA_{RG} does this. As in Section 1.2, suppose simple components defined by Rousson and Gasser are $W_k = X\gamma_k$, where the column means of X are 0 and the column lengths of X are 1. SCA_{RG} seeks a system of m simple components in b blocks maximizing the criterion $(Var(W_1) + \sum_{k=2}^{m} Var(W_k|W_1, \ldots, W_{k-1}))/\sum_{k=1}^{m} \lambda_k$, where λ_k $(k = 1, \ldots, m)$ is the eigenvalue of the k^{th} PC. In their paper, Rousson and Gasser call components whose non-zero loadings have all the same sign block components, and call components which have some positive and some negative loadings difference components. The definition of simplicity that Rousson and Gasser use has two factors. First, when there is more than one block component, in the case of an approximate block structure in the correlation matrix, small loadings may be ignored. Secondly, the weighting scheme of variables should be simple. All variables involved in a block component should have the same weight. For difference components, all positive weights should be equal, all negative weights be equal and the sum of all weights should be zero. This is similar to Chipman and Gu's contrast constraint.

The calculation of SCA_{RG} is divided into two stages (but SCA_{RG} is still defined as one stage method in this thesis because this method gets the simple components directly). The algorithm starts with a correlation matrix. In the first stage b simple block components are defined. In the second stage $q - b$ simple difference components are computed. Unfortunately, although the algorithm of Rousson and Gasser (2004) always gets simple components according to their definition, the components obtained by their methods are generally not orthogonal and not uncorrelated. For SCA_{RG}, block components are orthogonal to each other, and within-block difference components are orthogonal to block components but the difference components are not orthogonal to each other generally. Thus this algorithm loses two good properties of PCA.

I will give the results (Table 2.13) for Jeffers' pitprop data obtained using SCA_{RG} (Rousson and Gasser (2003)). As before only the first four components are retained. The interpretation of the components is straightforward. The first component is the average of the top diameter of the prop, the length of the prop, the maximum bow, the distance of the point of maximum bow from the top of the prop and the number of knot whorls. The second component is the average of the oven-dry specific gravity of the timber, the average diameter of the knots, the number of annual rings at the top and at the base of the prop. The third component denotes the average of the moisture content of the prop and the specific gravity of the timber. The fourth component represents the average number of knots per whorl. Comparing this results with the PCA in Table 1.9, the first four components by SCA_{RG} account for less variance (65%) than those of the first four principal components (74%) in Table 1.9, but the interpretation of each component is much easier than that of original PCs. This is expected because this type of method has to sacrifice some variance captured to improve interpretation of PCs. In Section 2.5, it was noted that the components obtained by SPCA are better than the components obtained by SCoTLASS and the truncation of loadings for the pitprop data, so we only compare the results in Table 2.13 with those in Table 2.12. The first four components in Tables 2.13 and 2.12 account for the same percent of variance (65% vs. 65%) but the components in Table 2.13 have fewer or the same number of non-zero loadings (5, 4, 4, 1 vs 7, 4, 4, 1), and the loadings of each component by SCA_{RG} just are the same for all the variables, so the interpretation of components by SCA_{RG} is much easier. Unfortunately, the angles of SCA_{RG} for pitprop data

Variables	Components			
	1	2	3	4
1-Topdiam	$1/\sqrt{5}$	0	0	0
2-Length	$1/\sqrt{5}$	0	0	0
3-Moist	0	0	$1/\sqrt{2}$	0
4-Testsg	0	0	$1/\sqrt{2}$	0
5-Ovensg	0	0.50	0	0
6-Ringtop	0	0.50	0	0
7-Ringbot	0	0.50	0	0
8-Bowmax	$1/\sqrt{5}$	0	0	0
9-Bowdist	$1/\sqrt{5}$	0	0	0
10-Whorls	$1/\sqrt{5}$	0	0	0
11-Clear	0	0	0	0
12-Knots	0	0	0	1
13-Diaknot	0	0.50	0	0
Number of nonzero loadings	5	4	2	1
Angle	35°	88°	70°	72°
Variance	25%	17%	15%	8%
Cumulative Variance	25%	42%	57%	65%

Table 2.13: SCA_{RG} for the pitprop data

are more than those obtained by SPCA (33°, 88°, 77°, 72° vs 28°, 44°, 44°, 37°). So the easier interpretation of the results by SCA_{RG} corresponds to bigger angles with the original PCs.

2.7 Discussion and conclusion

Chapter 2 reviews approaches that have been pursued for the simplification and interpretation of principal components. Rotation is a widely used technique in practice because it is readily available in software packages. However rotation has its drawbacks. For example rotated components are not invariant to a change in the number of components. A change from m to $m+1$ may change the nature of all the rotated PCs, because the rotation is now in $m+1$ space rather than in m space. Another common approach is to truncating loadings, this takes any loadings less than some threshold value as zero. Truncating loadings can be misleading (Cadima and Jolliffe, 1995). The variables with small loadings are not necessarily unimportant. Other approaches such as SCoTLASS, SCA_{RG}, SPCA etc are not popularly adopted in practice because these methods are too complicated to use or these methods are much newer so might not have had enough time to catch on. SCoTLASS, SCA_{RG} and SPCA just try to get some kind of modified components to enhance the interpretation of PCs without too much loss of variance. However the size of the angles between modified components and corresponding principal components is considered unimportant. The interpretation of the modified PCs produced by these three techniques is much easier than that of the corresponding PCs and the variance explained by modified PCs are usually slightly less than that of the PCs. Ignoring the size of angles generally makes the patterns of the modified components obtained by these three methods much different to those of the corresponding principal components.

The aims of all the techniques discussed in this chapter are to get components that are easy to interpret according to a definition of simplicity. Most of these techniques don't use an explicit definition of simplicity of the loadings, and don't care about the size of the angles between the modified PCs and original PCs. As was said in Section 1.2, PCs have two nice properties, they are orthogonal and uncorrelated. To enhance interpretation of PCs, techniques which modify PCs have to sacrifice at least one of these two properties. I prefer the techniques which have

explicit definition of simplicity and keep one of the two properties of PCA. Some techniques, such as that of Chipman and Gu (2003), have a clear definition of simplicity and consider the size of angles between the modified PCs and original PCs. Other techniques, such as that of Rousson and Gasser (2004), have a distinct definition of simplicity but do not consider the size of angles. Unfortunately, these two techniques lose both the two nice properties of PCs. This makes the measurement of the amount of variance each component accounts for difficult (Section 1.2). Vines (2000) provide the concept of simple principal components (SCA), the simple components are those that can be represented by integer vectors. As with PCA, the loading vectors of SCA are orthogonal with each other (keeping one of the best properties of PCA), but SCA achieves smaller-integer components that are easy to interpret, the results are obtained from the variance covariance matrix directly. So this is a one stage approach. More details of this technique will be discussed in the next chapter.

Chapter 3

The Simple Principal Components Algorithm

3.1 Introduction

In PCA, directions in the data are sought, for which the data have maximal variance. In general these directions are represented by loading vectors which can be difficult to interpret with respect to the original variables. Is it possible to find an algorithm which gives a good approximation of PCA but produces results that are easier to interpret than PCA? A new algorithm, simple principal components algorithm (SCA) (Vines, 2000), produces integer valued loadings vectors. These integer-valued loadings vectors approximate the loadings vectors obtained via PCA, but are hopefully more interpretable. This new algorithm works from the variance-covariance matrix directly. The algorithm consists of a series of simple linear transformations. Each transformation searches a restricted set of directions within a two-dimensional subspace for the directions with which the data have maximal variance.

Simple principal components algorithm will be discussed in this chapter. In Section 3.2, the classical Jacobi method, which is used to transform a real symmetric matrix to a diagonal matrix by a sequence of orthogonal transformations, is introduced. In Section 3.3, the SCA algorithm is described. In Section 3.4, six different types of SCA methods are provided. In Section 3.5,

SCA is applied to the same examples as Chapter 1, and the results from SCA are compared with those from the principal components analysis. In Section 3.6, the link between the SCA algorithm and the Jacobi algorithm is pinned down. In Section 3.7, some useful results about the difference (α) between SCA angle and Jacobi angle (θ) are obtained, and the relationship between angle α and θ is quantified. Section 3.8 is the discussion and conclusion.

3.2 Classical Jacobi transformations of a real symmetric matrix

The Jacobi method is used to transform a real symmetric $p \times p$ matrix A_0 to a diagonal matrix by a sequence of orthogonal transformations. The Jacobi method can be described as follows.

$$A_q = R_q^T A_{q-1} R_q \tag{3.1}$$

where A_q is the transformed matrix at iteration q. The transformation matrix R_q is determined by the following rules.

Suppose $a_{r,s}^{(q-1)}$ is the largest off-diagonal element of A_{q-1}, $r < s$. Then the rotation is chosen to be in the (r, s) plane with angle θ so that the $(r, s)th$ element of A_q becomes 0. The transformation matrix, R_q, is an identity matrix with just the four elements in positions (r, r), (s, s), (r, s) and (s, r) altered. For example, when $p = 6$, $r = 2$, $s = 5$ then

$$R_q = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \cos\theta & 0 & 0 & -\sin\theta & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \sin\theta & 0 & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The transformed matrix A_q is also symmetric. It differs from A_{q-1} only in rows r and s and columns r and s. Suppose $a_{ij}^{(q)}$ is the $(i, j)th$ element of matrix A_q. From equation (3.1), we have

$$a_{ir}^{(q)} = a_{ir}^{(q-1)} \cos\theta + a_{is}^{(q-1)} \sin\theta = a_{ri}^{(q)} \qquad (i \neq r, s) \tag{3.2}$$

$$a_{is}^{(q)} = -a_{ir}^{(q-1)} \sin\theta + a_{is}^{(q-1)} \cos\theta = a_{si}^{(q)} \qquad (i \neq r, s) \tag{3.3}$$

$$a_{rr}^{(q)} = a_{rr}^{(q-1)} \cos^2\theta + 2a_{rs}^{(q-1)} \cos\theta \sin\theta + a_{ss}^{(q-1)} \sin^2\theta \tag{3.4}$$

$$a_{ss}^{(q)} = a_{rr}^{(q-1)} \sin^2\theta - 2a_{rs}^{(q-1)} \cos\theta \sin\theta + a_{ss}^{(q-1)} \cos^2\theta \tag{3.5}$$

$$\begin{aligned} a_{rs}^{(q)} &= (a_{ss}^{(q-1)} - a_{rr}^{(q-1)}) \cos\theta \sin\theta + a_{rs}^{(q-1)} (\cos^2\theta - \sin^2\theta) \\ &= a_{sr}^{(q)} \end{aligned} \tag{3.6}$$

So setting $a_{rs}^{(q)}$ equal zero, if $a_{rr}^{(q-1)} \neq a_{ss}^{(q-1)}$ we get

$$\tan 2\theta_J = \frac{2a_{rs}^{(q-1)}}{a_{rr}^{(q-1)} - a_{ss}^{(q-1)}} \tag{3.7}$$

where θ_J lies in the range $|\theta_J| < \pi/4$.

If $a_{rr}^{(q-1)} = a_{ss}^{(q-1)}$ and $a_{rs}^{(q-1)} \neq 0$, θ_J is taken to be $\pi/4$ or $-\pi/4$ according to the sign of $a_{rs}^{(q-1)}$. Otherwise if $a_{rr}^{(q-1)} = a_{ss}^{(q-1)}$ and $a_{rs}^{(q-1)} = 0$, the matrix is diagonal already. So whatever the value of $a_{rr}^{(q-1)}$, $a_{ss}^{(q-1)}$ and $a_{rs}^{(q-1)}$, $|\theta_J| \leq \pi/4$.

From equations (3.2) and (3.3) we have

$$\sum_{i \neq r,s} [(a_{ir}^{(q)})^2 + (a_{is}^{(q)})^2] = \sum_{i \neq r,s} [(a_{ir}^{(q-1)})^2 + (a_{is}^{(q-1)})^2].$$

Hence the sum of the squares of the off-diagonal elements excluding the (r, s) and (s, r) elements is constant. Since a_{rs}^2 and a_{sr}^2 have been reduced by the rotation, overall the sum of the squared off-diagonal elements has reduced. Thus in this weak sense, matrix A is made closer to a diagonal matrix.

Adding equations (3.4) and (3.5), we get $a_{rr}^{(q)} + a_{ss}^{(q)} = a_{rr}^{(q-1)} + a_{ss}^{(q-1)}$. Because all the other diagonal elements at iteration q have the same values at iteration $(q-1)$, the sum of the diagonal elements at iteration q is the same as that at iteration $(q-1)$.

Now, suppose A_0 is a variance-covariance matrix, such that $a_{rr}^{(q-1)} \geq a_{ss}^{(q-1)}$ for every $r < s$. Otherwise, a transformation can be done to make the variance in position (r,r) greater than the variance in position (s,s). The transformation matrix is simply an identity matrix where the r^{th} and s^{th} columns have exchanged positions. So it is only necessary to consider this case. I am going to show that letting the largest off-diagonal element in position (r,s) become zero in each iteration is equivalent to finding the direction with maximum variance in position (r,r).

According to equation (3.4)

$$\begin{aligned} a_{rr}^{(q)} &= a_{rr}^{(q-1)} \cos^2\theta + 2a_{rs}^{(q-1)} \cos\theta \sin\theta + a_{ss}^{(q-1)} \sin^2\theta \\ &= a_{rr}^{(q-1)} \frac{(1+\cos 2\theta)}{2} + a_{rs}^{(q-1)} \sin 2\theta + a_{ss}^{(q-1)} \frac{(1-\cos 2\theta)}{2} \\ &= \left(\frac{a_{rr}^{(q-1)} - a_{ss}^{(q-1)}}{2} \right) \cos 2\theta + a_{rs}^{(q-1)} \sin 2\theta + \frac{(a_{rr}^{(q-1)} + a_{ss}^{(q-1)})}{2}. \end{aligned} \tag{3.8}$$

Because $a_{rr}^{(q-1)}$, $a_{ss}^{(q-1)}$ and $a_{rs}^{(q-1)}$ are constant, $a_{rr}^{(q)}$ is a function of θ. Supposing $a_{rr}^{(q)}$ is $f(\theta)$,

$$f'(\theta) = -(a_{rr}^{(q-1)} - a_{ss}^{(q-1)}) \sin 2\theta + 2a_{rs}^{(q-1)} \cos 2\theta.$$

and

$$f''(\theta) = -2(a_{rr}^{(q-1)} - a_{ss}^{(q-1)}) \cos 2\theta - 4a_{rs}^{(q-1)} \sin 2\theta$$

The angle θ^* with smallest absolute values which maximizes $a_{rr}^{(q)}$ is such that $f'(\theta^*) = 0$ and $f''(\theta) < 0$. If $a_{rr}^{(q-1)} = a_{ss}^{(q-1)}$ and $a_{rs}^{(q-1)} < 0$, $f'(\theta^*) = 2a_{rs}^{(q-1)} \cos 2\theta^* = 0$, $\theta^* = -\pi/4$, and $f''(\theta^*) = 4a_{rs}^{(q-1)} \sin 2\theta^* = 4a_{rs}^{(q-1)} < 0$. When $a_{rr}^{(q-1)} = a_{ss}^{(q-1)}$ and $a_{rs}^{(q-1)} > 0$, again $f'(\theta^*) = 0$, $\theta^* = \pi/4$, and $f''(\theta^*) = -4a_{rs}^{(q-1)} sin2\theta^* < 0$. So, when $a_{rr}^{(q-1)} = a_{ss}^{(q-1)}$, θ^* is the same angle as the Jacobi angle θ_J .

If $a_{rr}^{(q-1)} \neq a_{ss}^{(q-1)}$, let $f'(\theta) = 0$, we have

$$\tan 2\theta^* = \frac{2a_{rs}^{(q-1)}}{a_{rr}^{(q-1)} - a_{ss}^{(q-1)}}.$$

Again θ^* is the same as θ_J that results in $a_{rs}^{(q)} = 0$ (equation (3.7)) because $\theta_J < \pi/4$ when $a_{rr}^{(q-1)} \neq a_{ss}^{(q-1)}$, so $\theta^* < \pi/4$.

Next, I will prove that for $\theta = \theta^*$, $f(\theta)$ is maximized, i.e. $f''(\theta^*) < 0$. It is known that $\sin 2\theta^* = \frac{2a_{rs}^{(q-1)}}{a_{rr}^{(q-1)} - a_{ss}^{(q-1)}} \cos 2\theta$. So

$$f''(\theta^*) = -2(a_{rr}^{(q-1)} - a_{ss}^{(q-1)}) \cos 2\theta^* (1 + \tan^2 2\theta^*)$$

Under the hypothesis $a_{rr}^{(q-1)} > a_{ss}^{(q-1)}$, $|\theta^*| < \pi/4$, so that $\cos 2\theta^* > 0$, $f''(\theta^*) < 0$. So $f(\theta^*)$ still get the maximal value in this direction. All in all, θ^* is always equal to θ_J whatever the values of $a_{rr}^{(q-1)}$, $a_{ss}^{(q-1)}$ and $a_{rs}^{(q-1)}$, so $|\theta^*| \leq \pi/4$, $f(\theta^*)$ always get the maximal value in this direction. That is, Jacobi method seeking for the direction that results in $a_{rs}^{(q)} = 0$ is equivalent to seeking for the direction with maximal variance for element $a_{rr}^{(q-1)}$.

3.3 SCA algorithm

This section contains a description of SCA algorithm. It consists of four subsections. Section 3.3.1 talks about the linear transformation for two dimensional data at the heart of the SCA algorithm in each iteration. Section 3.3.2 gives the restriction of transformations in the SCA algorithm, so that the transformed directions are integer vectors if the directions in the previous iteration are integer vectors. Section 3.3.3 considers the comparison of transformations on different pairs of directions if the data has more than two directions. Section 3.3.4 gives a worked example.

3.3.1 Linear transformation of the SCA algorithm

The SCA algorithm consists of a series of orthogonal linear transformations, each transformation just affecting a two dimensional subspace in p-dimensional space. Suppose d_1 and d_2 represent these two directions. Assume that the lengths of d_1 and d_2 are 1 at iteration 0. In general the linear transformation is a rotation and rescaling. It can be written as $(f_1, f_2) = (d_1, d_2)B(\beta)$ where $B(\beta) = \begin{pmatrix} 1 & l_2^2\beta \\ \beta & -l_1^2 \end{pmatrix}$, β is a scalar and f_1 and f_2 are the transformed directions and l_1 and l_2 are the lengths of d_1 and d_2 respectively. That is

$$\left.\begin{aligned} f_1 &= d_1 + \beta d_2 \\ f_2 &= l_2^2 \beta d_1 - l_1^2 d_2 \end{aligned}\right\} \tag{3.9}$$

Now, when $d_1^T d_2 = 0$, in other words when d_1 and d_2 are orthogonal,

$$f_1^T f_2 = (d_1 + \beta d_2)^T (l_2^2 \beta d_1 - l_1^2 d_2) = l_2^2 \beta d_1^T d_1 - l_1^2 d_1^T d_2 + l_2^2 \beta^2 d_2^T d_1 - \beta l_1^2 d_2^T d_2 = 0.$$

So for all β, f_1 and f_2 are also orthogonal. Thus this transformation keeps orthogonality.

Again assuming $d_1^T d_2 = 0$, we also have

$$\begin{aligned} d_1^T f_1 &= d_1^T (d_1 + \beta d_2) = d_1^T d_1 = l_1^2 \\ d_2^T f_1 &= d_2^T (d_1 + \beta d_2) = d_2^T d_1 + \beta d_2^T d_2 = \beta l_2^2 \\ d_1^T f_2 &= d_1^T (l_2^2 \beta d_1 - l_1^2 d_2) = l_2^2 \beta d_1^T d_1 = \beta l_1^2 l_2^2 \\ d_2^T f_2 &= d_2^T (l_2^2 \beta d_1 - l_1^2 d_2) = -l_1^2 l_2^2 \end{aligned}$$

So

$$\begin{aligned} l_{f_1}^2 &= f_1^T f_1 = (d_1 + \beta d_2)^T f_1 = d_1^T f_1 + \beta d_2^T f_1 = l_1^2 + \beta^2 l_2^2 \\ l_{f_2}^2 &= f_2^T f_2 = (l_2^2 \beta d_1 - l_1^2 d_2)^T f_2 = l_2^2 \beta d_1^T f_2 - l_1^2 d_2^T f_2 = l_1^2 l_2^2 (l_1^2 + \beta^2 l_2^2) \end{aligned}$$

where l_{f_1} and l_{f_2} are the lengths of f_1 and f_2 respectively. So the lengths of f_1 and f_2 depend on the lengths of d_1 and d_2 and the value of β. Generally the length of f_1 is greater than the length of d_1 because $\beta^2 l_2^2 \geq 0$. This implies that f_1 is more complex than d_1 because generally the less simple components should have longer length . It is also known that $l_1^2 l_2^2 \geq 1$, so $1 \leq l_{f_1} \leq l_{f_2}$. This implies that generally f_1 will tend to be a more simple vector than f_2 and f_2 will tend to be more complex than d_2.

3.3.2 Restriction of transformations to directions in the algorithm

If the variance-covariance matrix corresponding the old axes d_1 and d_2 is V_0, then the variance-covariance matrix with respect to the new axes f_1 and f_2 is $B^T V B$. The variance v in the normalized direction f_1/l_{f_1} can be written as

$$v = \frac{l_1^2 v_{11} + 2\beta l_1 l_2 v_{12} + \beta^2 l_2^2 v_{22}}{l_1^2 + \beta^2 l_2^2}$$

where

$$V = \begin{pmatrix} l_1^2 v_{11} & l_1 l_2 v_{12} \\ l_1 l_2 v_{12} & l_2^2 v_{22} \end{pmatrix}$$

It can be shown that v is maximized by setting $\beta = \beta^*$

where $\beta^* = \begin{cases} \frac{-l_1 l_2 (v_{11} - v_{22}) + \sqrt{l_1^2 l_2^2 (v_{11} - v_{22})^2 + 4 l_1^2 l_2^2 v_{12}^2}}{2 l_2^2 v_{12}} & v_{12} \neq 0 \\ 0 & v_{12} = 0,\ v_{11} \geq v_{22} \\ \infty & v_{12} = 0,\ v_{11} < v_{22} \end{cases}$.

In this algorithm, the simple components are defined as integer components, i.e. the loading vectors are integer vectors. Rewriting f_1 as $f_1 = d_1 + \beta^* d_2$ and f_2 as $f_2 = l_2^2 \beta^* d_1 - l_1^2 d_2$, it can be seen that generally f_1 and f_2 are not integer vectors even if d_1 and d_2 are. This problem can be overcome by restricting β to be $b = \frac{i}{2^k}$ or $b = \frac{2^k}{i}$, $i = -2^k, -2^k + 1, \ldots, 2^k$ (Vines, 2000), where k is a nonnegative integer. The value of b which maximizes v is then sought only within this restricted set. Setting

$$\left.\begin{aligned} f_1 &= 2^k d_1 + 2^k b d_2 \\ f_2 &= 2^k b l_2^2 d_1 - 2^k l_1^2 d_2 \end{aligned}\right\} \quad |b| \leq 1 \tag{3.10}$$

$$\left.\begin{aligned} f_1 &= \frac{2^k d_1}{b} + 2^k d_2 \\ f_2 &= 2^k l_2^2 d_1 - \frac{2^k l_1^2 d_2}{b} \end{aligned}\right\} \quad |b| > 1 \tag{3.11}$$

ensures that f_1 and f_2 will be integer vectors whenever d_1 and d_2 are integer vectors. The transformation represented by equations (3.10) and (3.11) is called a simplicity preserving transformation. The simplicity preserving transformation matrix in equations (3.10) and (3.11) is $B(\beta)$ when $\beta = b$, taken as $B(b)$. $B(b)$ can be thought of as the transformation matrix in iteration 1. In iteration 2, the form of the transformation are the same as that in equations (3.10) and (3.11), the new directions are defined as d_1^2 and d_2^2, the old directions are f_1 and f_2, so the lengths of d_1 and d_2 in $B(b)$ become the lengths of f_1 and f_2. These process can be repeated until no

further non-trivial transformation exists. $B(b)_q$ represents the simplicity transformation matrix in iteration q. So the SCA algorithm consists of a sequence of simplicity preserving matrices $B(b)_q$. At each iteration q, v is maximized within the restricted set, as in the Jacobi method, the variance covariance matrix of SCA method in iteration q is $V_q = B(b)_q^T V_{q-1} B(b)_q$, V_q is more diagonal than V_{q-1}. Overall, the transformation matrix $B(b)$ is equal to $B(b)_1 B(b)_2 \ldots B(b)_n$ when the SCA method takes n iterations.

As has already been stated in this subsection, the transformed SCA directions in iteration 1 are controlled by the nonnegative integer k and the lengths of d_1 and d_2. For example when $k = 0$, f_1 is one of the four directions $d_1 - d_2$, d_1, $d_1 + d_2$, d_2 and f_2 is one of the four directions d_2, d_1 and $l_2^2 d_1 + l_1^2 d_2$, $l_2^2 d_1 - l_1^2 d_2$. If $l_1 = l_2$, $l_2^2 d_1 + l_1^2 d_2$ is the same as $d_1 + d_2$, and $l_2^2 d_1 - l_1^2 d_2$ is the same as $d_1 - d_2$. When $k = 1$, f_1 is one of the eight possible directions $d_1 - 2d_2$, $d_1 - d_2$, $d_1 - \frac{d_2}{2}$, d_1, $d_1 + \frac{d_2}{2}$, $d_1 + d_2$, $d_1 + 2d_2$ and d_2, f_2 is one of the eight directions d_1, d_2, $l_2^2 d_1 + l_1^2 d_2$, $l_2^2 d_1 - l_1^2 d_2$, $\frac{1}{2} l_2^2 d_1 + l_1^2 d_2$, $\frac{1}{2} l_2^2 d_1 - l_1^2 d_2$, $2 l_2^2 d_1 + l_1^2 d_2$, $2 l_2^2 d_1 - l_1^2 d_2$. When $l_1 = l_2$ the possible directions of f_1 and f_2 are the same. In general, when $l_1 \neq l_2$, some of the possible directions for f_1 and f_2 are the same and some are different.

The SCA directions available when $l_1/l_2 = 1$, 3 and 5 and $k = 0$, 1 and 2 are given in Figure 3.1. As k increases, the number of the directions available increases. The directions for larger k include all the directions for small k. For the same k, the number of directions when $l_1/l_2 = 3$ and $l_1/l_2 = 5$ are the same, and more than that when $l_1 = l_2$.

3.3.3 Comparison of transformations on different pairs of directions

So far, the SCA algorithm has been discussed assuming that the data has just two directions. If the data has more than two directions, how does the SCA algorithm work? Assume now the data is described with respect to p orthogonal simple directions d_1, d_2, ... d_p, all of their lengths are 1 and directions r and s are updated in iteration q. The variance-covariance matrix in iteration q is V_q. The 2 by 2 simplicity preserving transformation matrix $B(b)_q$ as described in Section 3.3.2 is extended to a p by p matrix P_q. Then, $V_q = P_q^T V_{q-1} P_q$, $q = 1, 2, \ldots$. Also

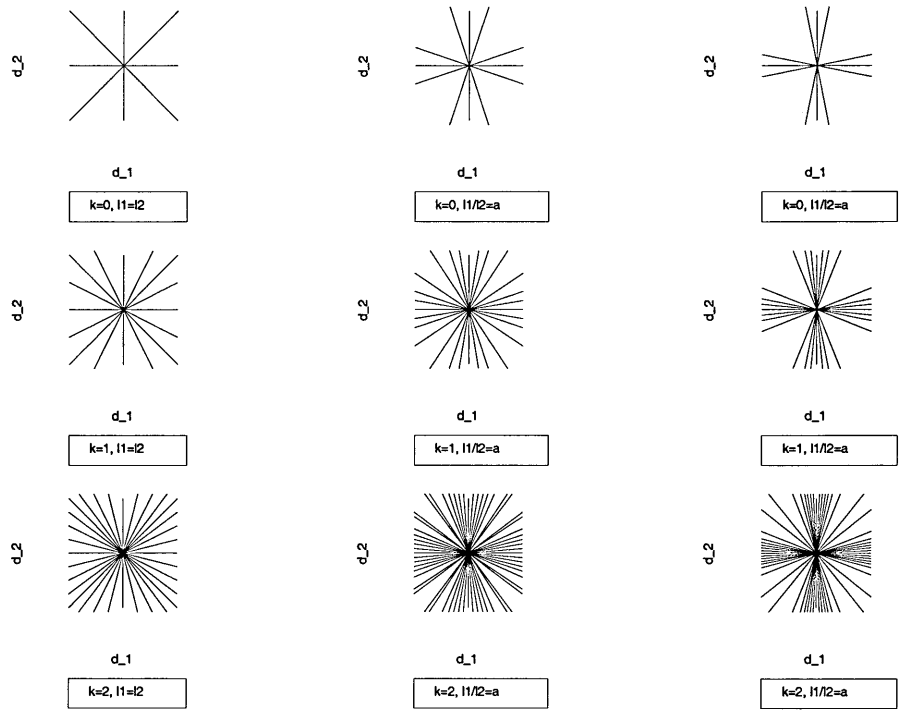

Figure 3.1: The directions available for SCA methods when $k = 0$, $k = 1$ and $k = 2$ and l_1/l_2=1, 3, 5

the matrix of the directions after the q^{th} iteration, $D_q = (d_1^q, \ldots, d_p^q)$, is $D_q = D_{q-1}P_q$. As with two directions, each transformation P_q results in the variance-covariance matrix becoming more diagonal. The process ends when no non-trivial pairwise transformation, P_q, is indicated. The overall transformation matrix $P = P_1P_2 \ldots P_n$. For a 2 by 2 simplicity preserving matrix, there are only two directions (one pair), these two directions are updated in every iteration. For a p by p simplicity preserving matrix, each transformation matrix P_q, corresponds to a single pairwise transformation (single updating) or to more than one pairwise transformation (multiple updating), each involving different pairs of directions. For single updating, at every iteration, only one pair is updated. The simplicity preserving transformation matrix P_q is a $p \times p$ identity matrix with the four elements in positions (r, r), (r, s), (s, s) and (s, r) altered. These four elements are the same as those in the transformation matrix represented by equations (3.10)

and (3.11). For example, when $|b| \leq 1$, $p = 6$, $r = 2$, $s = 5$, l_2 and l_5 represent the lengths of d_2 and d_5 in iteration $q-1$. Then the simplicity transformation matrix for single updating in iteration q is

$$P_q = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2^k & 0 & 0 & 2^k b l_5^2 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 2^k b & 0 & 0 & -2^k l_2^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

So P_q depends on the nonnegative integer k, b and the length of updated directions in the previous iteration.

For single updating, if the updating directions in iteration q are known, the form of P_q is determined. So how to choose the updating directions (the order in which simplicity preserving transformations are performed) for single updating methods at iteration q? Various ways can be used to choose the order in which the simplicity preserving transformations are performed. The relative merit of the pairwise transformations can be judged either by the maximal variance obtained in one of the directions (maximal variance criterion) or by the maximal improvement in variance (maximal improvement in variance criterion) as a result of the transformation. The maximal variance criterion means that the pair of directions r and s is chosen so that $v_q(r,r)$ is maximized. The maximal improvement in variance criterion chooses, at iteration q, the transformation for which $(v_q(r,r) - v_{q-1}(r,r))$, $r = 1, \ldots, p$, is the greatest.

For multiple updating, at iteration q, the first pair is chosen in the same way as single updating, i.e. the pair with maximal variance or maximal improvement in variance is chosen. However, the columns (updating directions) involved in the next transformation are restricted to the pair with maximal variance or maximal improvement in variance except for the columns updated in the previous iteration. This process continues until there are no further non-trivial transformations that can be carried out in iteration q. Generally n pairs of the directions (where $n \leq p/2$) are updated simultaneously at each iteration. So the P_q for multiple updating equals the multiplication of n single updating matrices. The order in which these single updating

r s	max($v_0(r,r)$, $v_0(s,s)$)×100	$v_1(r,r)$ ×100	Improvement ×100	b value
1 2	1.55	2.24	0.69	b=1
1 3	2.54	3.31	0.77	b=1
1 4	2.66	2.78	0.12	b=1
2 3	2.54	2.66	0.12	b=1
2 4	2.66	3.39	0.73	b=1
3 4	2.66	3.36	0.70	b=1

Table 3.1: The maximal variance and value b for the RI data in iteration 1

matrices are multiplied together is unimportant because each single updating matrix updates different pairs by construction.

Finally an extra constraint can be imposed on the pairwise transformations, limiting consideration to pairwise transformations of directions of the same length.

So, in general, the SCA algorithm is divided into five stages: (i) start with a p by p variance covariance or correlation matrix, (ii) choose the updating directions according to the criterion (maximal variance or maximal improvement in variance), (iii) the transformation matrix P_q is determined according to single updating or multiple updating, (iv) repeat stages 2 and 3, until no further non-trivial transformation exists, (v) the overall transformation matrix P is equal to $P_1P_2 \ldots P_n$. When n is the number of iterations the SCA method takes.

As has already been said, single updating only updates one pair at each iteration whereas multiple updating updates as many pairs as possible in each iteration. In this thesis, updating one pair is defined as one step. So for single updating, one iteration always consists of just one step. For multiple updating, one iteration generally consists of more than one step.

3.3.4 The transformation matrix of RI data in iteration 1, $k = 0$

A comparison of the first transformation on different pairs of directions are demonstrated for the RI data.

Consider the results in Table 3.1, the first column is the maximal variance (multiplied by 100) in positions (r,r) and (s,s) at iteration 0. For example, at iteration 0

$$V_0 = \begin{pmatrix} 1.55 & 0.70 & 1.26 & 0.67 \\ 0.70 & 1.53 & 0.63 & 1.30 \\ 1.26 & 0.63 & 2.54 & 0.76 \\ 0.67 & 1.30 & 0.76 & 2.66 \end{pmatrix}.$$

So the variances of directions 1 and 2 are 1.55 and 1.53 respectively, and hence the maximum variance is 1.55. The second column is the maximum variance multiplied by 100 in position (r,r) after columns r and s are updated in iteration 1 for different b. When $k=0$, the possible values of b are 0 and ± 1. For example, after directions 1 and 2 are updated, $v_1(1,1)$ is 1.55 when $b=0$, $v_1(1,1)$ is 0.84 when $b=-1$, $v_1(1,1)$ is 2.24 when $b=1$. So the best b is 1. The third column is the maximal improvement in variance in position (r,r) after directions r and s are updated in iteration 1. Each value in the improvement in variance is equal to $v_q(r,r) - v_{q-1}(r,r)$, i.e. the value in column 2 minus the value in column 1 is equals to the maximal improvement in variance after directions r and s are updated. For example, when $r=1$, $s=2$, the maximal improvement in variance is $2.24 - 1.55 = 0.69$, again the best b is 1. In the first iteration, the maximal variance and the maximal improvement in variance are obtained when $b=1$ for every pair of variables. The maximal variance (3.39) over all pairs r and s is obtained after updating the directions 2 and 4, the next pair with maximal variance which does not include directions 2 and 4 is $(1,3)$. In contrast, the maximal improvement in variance is 0.77, obtained after updating the directions 1 and 3. The next pair ignoring directions 1 and 3 is directions 2 and 4. There is improvement in variance (0.73) after updating directions 2 and 4.

So if single updating and maximal variance criterion are used, directions 2 and 4 are updated in the first iteration and $b=1$, hence

$$P_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}.$$

If single updating and maximal improvement in variance are used, directions 1 and 3 are updated in the first iteration and $b = 1$, So

$$P_1 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

When multiple updating and the maximal variance criterion are used, as stated before, directions $(2,4)$ and $(1,3)$ are updated at the same time in iteration 1. Similarly, if the criterion is maximal improvement in variance, the best choice of the directions is $(1,3)$, followed by $(2,4)$. So P_1 is the same whether the multiple updating with maximal variance criterion is used or the maximal improvement variance criterion is used. This is because for multiple updating, in the first iteration for the RI data, the pairs selected with maximal variance and maximal improvement in variance criteria are the same, and the transformation matrices are the same. So the variance covariance matrices at iteration 2 are the same. However generally the transformation matrices and the resultant variance covariance matrices differ with different criteria. For example, as has been stated in this subsection, for single updating, in the second iteration, the variance covariance matrices are different for different criteria because P_1 is different for different criteria. Thus

$$P_1 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}.$$

3.4 Types of SCA methods

In this section, six different SCA methods will be introduced. These methods will be investigated in the following chapters. They are called SCA1 to SCA6 respectively.

In each iteration, SCA1, SCA2 and SCA5 use the maximal improvement in variance criteria, they are called SCA methods with maximal improvement in variance. In contrast, SCA3, SCA4

Method	Criteria	Updating	$l_r = l_s$
1	Maximal improvement in variance	Single	Yes
2	Maximal improvement in variance	Single	No
3	Maximal variance	Single	Yes
4	Maximal variance	Single	No
5	Maximal improvement in variance	Multiple	No
6	Maximal variance	Multiple	No

Table 3.2: SCA methods used in the rest of the thesis

and SCA6 use the maximal variance criterion, they are called SCA methods with maximal variance. SCA1, SCA2, SCA3 and SCA4 update one pair at each iteration (single updating, Section 3.3.3), they are called single methods. SCA5 and SCA6 update as many pairs as possible at each iteration (multiple updating, Section 3.3.3), and are called multiple methods.

SCA1 and SCA2 are the single SCA methods with maximal improvement in variance. The only difference is that at every iteration the extra constraint $l_r = l_s$ is imposed on SCA1. SCA3 and SCA4 are the single SCA methods with maximal variance. The only difference is that at every iteration the extra constraint $l_r = l_s$ is imposed on SCA3. The extra condition $l_r = l_s$ is imposed on the SCA methods because if the lengths of all the components produced by SCA methods are equal the components produced by SCA methods are generally simple. All the SCA methods are listed in Table 3.2.

3.5 Examples

In this section, the SCA methods will be applied to the same four examples given in Chapter 1. In the first two examples, the results obtained using SCA5 (multiple SCA method with maximal improvement in variance) with $k = 0$, 1 and 2 are given. In the third example I only consider the results using SCA5 when $k = 0$ and in the last example I consider the results using SCA5 and SCA2 (single SCA method with maximal improvement in variance) when $k = 0$.

3.5.1 Example 1 - RI data

The RI data was introduced in Section 1.3.1. As with PCA given in Section 1.3.1, the SCA method starts with the variance-covariance matrix. The original variance covariance matrix (multiplied by 100) is V_0 as given in Table 1.1,

$$V_0 = \begin{pmatrix} 1.55 & 0.70 & 1.26 & 0.67 \\ 0.70 & 1.53 & 0.63 & 1.30 \\ 1.26 & 0.63 & 2.54 & 0.76 \\ 0.67 & 1.30 & 0.76 & 2.66 \end{pmatrix}.$$

d_1, d_2, d_3 and d_4 are the original four directions vectors. Recall from Section 3.3.3 that for multiple updating, the first transformation matrix, P_1, is

$$P_1 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}.$$

So this first iteration has two steps.

The new directions in iteration 1 $(d_1^1, d_2^1, d_3^1, d_4^1) = (d_1, d_2, d_3, d_4)P_1$, the lengths of d_1^1, d_2^1, d_3^1 and d_4^1 are $\sqrt{2}$, and the variance covariance matrix corresponding to d_1^1, d_2^1, d_3^1 and d_4^1 in the first iteration is $P_1^T V_0 P_1$. So the variance covariance at iteration 1 corresponding to the normalized directions d_1^1/(length of d_1^1), d_2^1/(length of d_2^1), d_3^1/(length of d_3^1), and d_4^1/(length of d_4^1) is

$$V_1 = \begin{pmatrix} 3.31 & 1.38 & -0.50 & -0.06 \\ 1.38 & 3.39 & -0.01 & -0.57 \\ -0.50 & -0.01 & 0.79 & 0.08 \\ -0.06 & -0.57 & 0.08 & 0.80 \end{pmatrix}.$$

After the first transformation, the off diagonal elements of V_1 generally are less than those of V_0, so V_1 is more diagonal than V_0. In the second iteration, the maximal variance, the maximal improvement in variance and the corresponding value of b are listed in the Table 3.3.

Comparing the transformations (as given in Table 3.3) involving the different pairs of directions, the best transformation using maximal improvement in variance is the one involving

r	s	$\max(v_1(r,r), v_1(s,s)) \times 100$	$v_2(r,r) \times 100$	Improvement $\times 100$	b value
1	2	3.39	4.73	1.34	b=1
1	3	3.31	3.31	0	b=0
1	4	3.31	3.31	0	b=0
2	3	3.39	3.39	0	b=0
2	4	3.39	3.39	0	b=0
3	4	0.80	0.88	0.08	b=1

Table 3.3: The maximal variance and value b for the RI data at iteration 2

d_1^1 and d_2^1. The transformation including the remaining two directions d_3 and d_4 also has an improvement in variance. Thus it is performed simultaneously. So this second iteration also has two steps and the second transformation matrix is

$$P_2 = \begin{pmatrix} 1 & -2 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 1 & 2 \end{pmatrix}.$$

The updated covariance matrix, V_2, is

$$V_2 = \begin{pmatrix} 4.73 & 0.04 & 0.56 & 0.06 \\ 0.04 & 1.97 & 0.01 & 0.50 \\ 0.56 & 0.01 & 0.87 & 0.00 \\ 0.06 & 0.50 & 0.00 & 0.72 \end{pmatrix}.$$

Again, V_2 is more diagonal than V_1.

At the third iteration there are no more non-trivial simplicity preserving transformations that can be applied, so the algorithm is over.

The overall transformation is $P = P_1 P_2$, so that

r	s	$\max(v_1(r,r), v_1(s,s)) \times 100$	$v_2(r,r) \times 100$	Improvement $\times 100$	b value
1	2	3.39	4.73	1.34	b=1
1	3	3.31	3.31	0	b=0
1	4	3.31	3.31	0	b=0
2	3	3.39	3.39	0	b=0
2	4	3.39	3.39	0	b=0
3	4	0.80	0.88	0.08	b=1

Table 3.3: The maximal variance and value b for the RI data at iteration 2

d_1^1 and d_2^1. The transformation including the remaining two directions d_3 and d_4 also has an improvement in variance. Thus it is performed simultaneously. So this second iteration also has two steps and the second transformation matrix is

$$P_2 = \begin{pmatrix} 1 & -2 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 1 & 2 \end{pmatrix}.$$

The updated covariance matrix, V_2, is

$$V_2 = \begin{pmatrix} 4.73 & 0.04 & 0.56 & 0.06 \\ 0.04 & 1.97 & 0.01 & 0.50 \\ 0.56 & 0.01 & 0.87 & 0.00 \\ 0.06 & 0.50 & 0.00 & 0.72 \end{pmatrix}.$$

Again, V_2 is more diagonal than V_1.

At the third iteration there are no more non-trivial simplicity preserving transformations that can be applied, so the algorithm is over.

The overall transformation is $P = P_1 P_2$, so that

r	s	$\max(v_1(r,r), v_1(s,s)) \times 100$	$v_2(r,r) \times 100$	Improvement $\times 100$	b value
1	2	4.73	4.73	0	b=0
1	3	4.73	4.73	0	b=0
1	4	4.73	4.73	0	b=0
2	3	1.97	1.97	0	b=0
2	4	1.97	1.97	0	b=0
3	4	0.87	0.87	0	b=0

Table 3.4: The maximal variance and value b for the RI data at iteration 3

$$P = \begin{pmatrix} 1 & -2 & -1 & 2 \\ 1 & 2 & -1 & -2 \\ 1 & -2 & 1 & -2 \\ 1 & 2 & 1 & 2 \end{pmatrix}.$$

The simple components are proportional to the columns of matrix P. Note that the second and fourth components can be divided through by 2 and still be integers. The simple components are orthogonal by construction, so the total variance of the four variables explained by SCA methods are the same as that explained by PCA. So the variance each simple component explains is equal to the variance of each component divided by the total variance of the four variables. It is shown in Table 3.5 that the simple components explain 57%, 24%, 11% and 9% of the variance respectively. The first two components together explain 81% of the total variance. Comparing this with the PCs in Table 1.2, the variances obtained by the components produced by SCA5 are very similar with the variance of the corresponding components produced by PCA (57% vs 58%, 24% vs 26%, 11% vs 10%, 9% vs 7%). Also the angles between the simple components and the original PCs are only 8.3°, −19.3°, −8.2° and 19.3° respectively. So the approximation of the components producing by SCA to principal components is very good for the RI data.

Furthermore, the interpretation of the simple components is straightforward. All the absolute value of the loadings of simple components are equal to 1. So the interpretation of the

		Component			
		1	2	3	4
Variables	Dop-R	1	1	1	1
	Dop-L	1	-1	1	-1
	Cvi-R	1	1	-1	-1
	Cvi-L	1	-1	-1	1
	Angle	8.3°	-19.3°	-8.2°	19.3°
	Variance	4.73	1.97	0.87	0.72
	Cumulative variance	57%	81%	92%	100%

Table 3.5: Simple components for the RI data

simple components are much easier than the PCs for RI data in Section 1.3.1. In particular the interpretation of components 2 and 4 is easier because all of the loadings of the second and fourth PCs are intermediate in value. The first simple component represents an overall measurement of the RI. It's just the average of all the variables. The second component represents a contrast between the RI in uterine arteries on the right and left sides of the body. The third component represents a contrast between measurements of the RI using the two different techniques. One technique has loadings of $+1$, the other technique has loadings of -1. Finally the fourth component represents the interaction between location and technique.

Table 3.6 shows that the higher the k for RI data, the better the approximation because of the smaller angles but the more difficult the interpretation. For this example, the results when $k = 0$ give the best balance between approximation and interpretation.

3.5.2 Example 2 - Sparrow data

Consider the sparrow data set introduced in Section 1.3.2. As in Section 1.3.2, start with the correlation matrix V_0 as given in Table 1.3,

	Component	Angle	Directions			
k=0	1	8.3°	1	1	1	1
	2	-19.3°	1	-1	1	-1
	3	-8.2°	1	1	-1	-1
	4	19.3°	1	-1	-1	1
k=1	1	10.4°	1	1	2	2
	2	-1.6°	1	-1	2	-2
	3	-10.4°	2	2	-1	-1
	4	-1.7°	-2	2	1	-1
k=2	1	1.4°	3	3	4	4
	2	-3.2°	8	-8	19	-19
	3	1.4°	-4	-4	3	3
	4	3.2°	19	-19	-8	8

Table 3.6: Simple components for the RI data

		Component				
		1	2	3	4	5
Variables	Total length	1	0	-1	-1	-3
	Alar extent	1	-1	-1	1	2
	Length of beak and head	1	-1	1	-1	2
	Length of humerus	1	0	1	1	-3
	Length of keel of sternum	1	2	0	0	2
	Angle	3.3°	13.9°	18.7°	11.2°	18.1°
	Variance	3.61	0.52	0.37	0.31	0.19
	Cumulative variance	72%	82%	90%	96%	100%

Table 3.7: Simple components for the sparrow data

$$V_0 = \begin{pmatrix} 1.00 & 0.74 & 0.66 & 0.65 & 0.61 \\ 0.74 & 1.00 & 0.67 & 0.77 & 0.53 \\ 0.66 & 0.67 & 1.00 & 0.76 & 0.53 \\ 0.65 & 0.77 & 0.76 & 1.00 & 0.61 \\ 0.61 & 0.53 & 0.53 & 0.61 & 1.00 \end{pmatrix}.$$

In this example, the simple components produced by using SCA5 are as given in Table 3.7. The simple components explain 72%, 10%, 8%, 6% and 4% of the variance respectively. So the first two components together explain 82% of the total variance. The variance and the percentage of total variance each component accounts for are very similar to the results of PCA given in Table 1.4 (72% vs 72%, 10% vs 11%, 8% vs 8%, 6% vs 6%, 4% vs 3%). The angles between the simple components and the original PCs are only 3.3°, 13.9°, 18.7°, 11.2° and 18.1° respectively. So the approximation of the SCA are quite good to PCA for Sparrow data.

The interpretation of the components is straightforward. The first simple component represents an overall measurement of the sparrows. All of the loadings of the first simple component are equal to 1, it is just the average of all the five standardised variables, where as all of the

loadings of first PC are similar but not the same. The second simple component represents a contrast between the alar extent, length of beak and head and twice the length of keel of sternum. The last variable seems important for this component because the loading of the last variable is 2. That is to say, the second component is high if the length of keel of sternum is high, and low if alar extent and the length of beak and head are low. It appears to represent a shape difference in the sparrows. Comparing the second simple component with the second PC in Section 1.3.2, there are two exact zero loadings for the simple component, but only one loading near 0 for the second PC. The non-zero loadings of the second SC are -1 or 2 respectvely. Most of the non-zero loadings of the second PC are intermediate in value. The third component just represents the contrast between total length, alar extent and length of beak and head, length of humerus because the loadings of the first two variables are just -1 and the loadings of the third and fourth variables are just $+1$. The last loading for the third simple component is 0, the last loading for the third PC are not zero, but for the other loadings, SC loadings are $+1$ or -1, it was very easy to interpret. Similarly, the absolute loadings for the fourth SC are 1 except the last loading (for the fourth PC, this loading is very small). The fourth component is the contrast between total length, length of beak and head and alar extent, length of humerus. Finally the fifth component represents the interaction between total length, length of humerus and alar extent, length of beak and head and length of keel of sternum. So, the third, fourth and fifth components represent other aspects of shape difference. The interpretation of simple components are much easier than the PCs for Sparrow data in Section 1.3.2 because of the integer loadings.

The bigger the k, the smaller the angles (Table 3.8), but the components are not really simple for large k because the loadings of the simple components are too big. So for this data set, $k = 0$ is the best choice. Comparing the simple components of sparrow data when $k = 0$ with the principal components, the approximation is good.

3.5.3 Example 3 - Employment in Europe data

For the RI data and Sparrow data, $k = 0$ gave the best choice, the best balance between approximation and interpretation. So in the following two examples, I only consider the results

	Comp.	Var.	Ang.	Directions				
k=0	1	3.61	3.3°	1	1	1	1	1
	2	0.52	13.9°	0	-1	-1	0	2
	3	0.37	18.7°	-1	-1	1	1	0
	4	0.31	11.2°	-1	1	-1	1	0
	5	0.19	18.1°	-3	2	2	-3	2
k=1	1	3.61	3.3°	1	1	1	1	1
	2	0.53	10.7°	3	-7	-9	-3	16
	3	0.39	-7.9°	230	101	-116	-127	-88
	4	0.30	-5.9°	3209	-5290	6153	-3888	-184
	5	0.17	9.1°	-5058919	7079297	3067665	-9144711	4056668
k=2	1	3.61	3.0°	4	4	4	4	3
	2	0.54	-3.5°	-12	79	80	45	-256
	3	0.39	-2.3°	5439	2499	-3769	-3245	-1232
	4	0.30	3.3°	-79731	106427	-124619	88083	13120
	5	0.17	3.4°	-9266649	13779378	7420129	-15184978	4336160

Table 3.8: Simple components for the sparrow data

Variables	Component 1	2	3
1-Agriculture	-1	0	-1
2-Mining	0	1	-1
3-Manufacturing	1	0	0
4-Power supplies	1	0	-2
5-Construction	1	0	0
6-Service industries	0	-1	0
7-Finance	0	-1	-1
8-Social and personal services	1	0	1
9-Transport and communications	1	0	0
Angles	25.7°	34.8°	-39.7°
Variance	3.00	1.80	1.07
Cumulative variance	33%	53%	65%

Table 3.9: Simple components for the employment in Europe data

when $k = 0$. As in Section 1.3.3, for the employment in Europe data only the first three components obtained by SCA5 are considered.

The first three simple components explain 33%, 20% and 12% of the variance respectively (Table 3.9). The interpretation of the first few components is straightforward. There are 3 zero loadings for first SC but only 2 near zero loadings for first PC, one variable, service industries, which has a zero loading for the first SC but has loading 0.38 which is the third biggest loading for first PC. The signs of the non-zero loadings for first SC is the same as the signs of corresponding loadings of first PC, but the absolute value of the loadings are all 1 for the first SC. So the interpretation of the first SC are different and much easier than that of the first PC. The first component represents the difference between the percentage of labor force in agriculture and that in manufacturing, power supplier, construction, social and personal service, transport and

communication. The second SC is particularly simple, there are 6 zero loadings and the absolute values of the other loadings are 1 but for second PC only two loadings near 0 and other loadings take intermediate values. The second component is the difference between mining and service industries, finance. The second SC and second PC have quite different interpretation. For the similar reason as the second SC, the interpretation of the third SC is easier than that of the third PC. The third component is the difference between social and personal services and agriculture, mining, power supplies, finance.

The variance explained for each simple component is less than that for each PC (3.00 vs 3.49, 1.80 vs 2.13, 1.07 vs 1.10). So the variance explained by the first three simple components is only 65% compared to 75% for the PCs in Section 1.3.3. The angles between the simple components and the original PCs are 25.7°, 34.8° and −39.7° respectively. This means that the approximation for this data is not so good as that for RI data and Sparrow data. Furthermore in the PCA solution, the final component has a variance of 0, the SCA solution fails to pick this up. Instead the final component of the SCA solution has a variance of 0.23. However this results are much better than the rotated results for this data in Section 2.2 because of the easier interpretation of the components and the smaller angles.

3.5.4 Example 4 - Jeffers' pitprop data

As was said in previous subsection, for this example only consider the results when $k = 0$. As in Section 1.3.4, only the first 4 components are retained.

The simple components produced by SCA5 for the pitprop data when $k = 0$ are not as close to corresponding PCs as in the previous three examples (Table 3.9). Furthermore, the loadings of the fourth component are too big to be simple and interpretable. The simple components produced with the higher k are also disappointing. So consider other SCA methods when $k = 0$. It is shown that the angles between PCs and the corresponding components produced by SCA2 are the smallest except for component 3 (the angle for component 3 is not too big), so the simple components produced by SCA2 are more close to the corresponding PCs than those produced by other SCA methods, and the components produced by SCA2 are easier to interpret. So the simple components produced by SCA2 gave the best balance between approximation

	Component			
Variables	1	2	3	4
1-Topdiam	1	2	2	-133
2-Length	1	2	2	-133
3-Moist	0	1	-9	603
4-Testsg	0	1	-9	601
5-Ovensg	1	-2	-5	79
6-Ringtop	1	0	-11	-333
7-Ringbot	1	0	-9	-273
8-Bowmax	1	0	8	250
9-Bowdist	1	1	1	-68
10-Whorls	1	0	8	224
11-Clear	0	1	0	20
12-Knots	-1	1	-7	-308
13-Diakno	-1	2	3	-79
Angles	27.4°	43.4°	-34.3°	73.9°
Variance	3.67	2.40	1.81	1.05
Cumulative variance	28%	47%	61%	69%

Table 3.10: Simple components obtained by SCA5 for the Jeffer's pitprop data

	Component			
Variables	1	2	3	4
1-Topdiam	1	1	1	0
2-Length	1	1	1	0
3-Moist	0	1	-3	0
4-Testsg	0	1	-3	0
5-Ovensg	0	-1	-1	0
6-Ringtop	1	-1	-1	0
7-Ringbot	1	-1	-1	0
8-Bowmax	1	0	0	0
9-Bowdist	1	0	0	0
10-Whorls	1	0	0	0
11-Clear	0	0	0	1
12-Knots	0	0	0	0
13-Diakno	0	1	1	0
Angles	-17.5°	42.6°	-45.1°	36.5°
Variance	3.92	1.97	1.69	1.00
Cumulative variance	30%	45%	58%	66%

Table 3.11: Simple components obtained by SCA2 for the Jeffers' pitprop data

and interpretation amongst all the SCA methods. The results obtained by SCA2 are given in Table 3.10. Table 3.11 shows that the first four simple components explain 30%, 15%, 13% and 8% of the variance respectively. The interpretation of the components is straightforward. The total variance explained by the first four components is 66%, this is slightly less than the total variance explained by the corresponding components obtained by SCA5 (69%). However the simple components obtained by SCA2 is much easier to interpret than the simple components obtained by SCA5. This is because each component obtained by SCA2 has more zero loadings than corresponding component obtained by SCA5, and the non-zero loadings are 1 or very small integers. Furthermore, the simple components obtained by SCA2 are much easier to interpret than PCs in Section 1.3.4. For the first SC there are 6 zero loadings and the other 7 loadings are 1 . In contrast for the first PC most of the loadings are intermediate in value. The first simple component represents an overall measurement of the top diameter, length, the number of rings at the top and base of the prop, the bow and the number of whorls. It is just the average of these variables. Similarly, the second component represents a difference between the top diameter, length, the moisture content, the specific gravity, the average diameter of the knots and the oven-dry specific gravity, the number of rings at the top and base. The equal absolute non-zero loadings of the second PC and 5 exact zero loadings make the interpretation of SC easier. The third component represents a difference between the top diameter, length, the average diameter of the knots and the moisture content, the specific gravity, the oven-dry specific gravity, the number of rings at top and base of the prop. The fourth component represents a measure of the length of clear prop from the top.

Comparing the simple components with principal components, the variance explained by each simple component is less than that of PCA (3.92 vs 4.22, 1.97 vs 2.38, 1.69 vs 1.88, 1.00 vs 1.11). So the variance explained by the first four simple components reduces to 66% compared to 74% for the PCs. The angles between the simple components and the original PCs are $-17.5°$, $42.6°$, $-45.1°$ and $36.5°$ respectively. So the approximation of the PCs by the SCs for the pitprop data is not so good as that for the RI data and the Sparrow data. However, the angles here are less than those of SCoTLASS, SCA_{RG} and SPCA in Chapter 2. So overall the SCA solution is more close to PCs than that achieved by the other methods for simplifying PCs for this data.

3.6 SCA algorithm links with Jacobi method

As was said in Sections 3.2 and 3.3, each step in the SCA algorithm and in the Jacobi method affects two directions. So we consider here the link of the SCA algorithm with the Jacobi method with two directions d_1 and d_2. In this section I am going to show that the linear transformation of the SCA method in each step is equivalent to rescaling d_1 and d_2, orthogonally rotating by an angle (the SCA angle), rescaling f_1 and f_2, and finally reflecting d_1 and d_2. Suppose the two methods start from a 2×2 variance-covariance matrix $V_0 = \begin{pmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{pmatrix}$. where V_0 is the variance-covariance matrix with respect to the normalized directions $\frac{d_1}{l_1}$ and $\frac{d_2}{l_2}$. Also suppose that $v_{11} \geq v_{22}$ and v_{12} is not equal to 0, i.e. that V_0 is not already diagonal. The transformation matrix of the two dimensional Jacobi method is

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Following equation (3.7), $\tan 2\theta = \frac{2v_{12}}{v_{11}-v_{22}}$. Now consider the SCA method. As shown in Section 3.3.1, in general the linear transformations at the heart of SCA can be written as $(f_1, f_2) = (d_1, d_2)B(b)$ where b is a scalar, f_1 and f_2 are the transformed directions, and $B(b) = \begin{pmatrix} 1 & l_2^2 b \\ b & -l_1^2 \end{pmatrix}$. Suppose the rotating angle of the SCA algorithm is $\theta + \alpha$, θ is the Jacobi angle, α is the difference between SCA angle and Jacobi angle. R_1 is the rotation matrix of the SCA method. According to the results obtained in Section 3.3.1,

$$\cos(\theta + \alpha) = \frac{d_1^T f_1}{l_1 l_{f_1}} = \frac{l_1}{\sqrt{l_1^2 + b^2 l_2^2}} \tag{3.12}$$

$$\sin(\theta + \alpha) = \frac{d_2^T f_1}{l_2 l_{f_1}} = \frac{b l_2}{\sqrt{l_1^2 + b^2 l_2^2}} \tag{3.13}$$

It is known from equation (3.12) that $\cos(\theta + \alpha) > 0$, thus $|\theta + \alpha| < \frac{\pi}{2}$. Thus

$$R_1 = \begin{pmatrix} \cos(\theta + \alpha) & -\sin(\theta + \alpha) \\ \sin(\theta + \alpha) & \cos(\theta + \alpha) \end{pmatrix}.$$

That is

$$R_1 = \begin{pmatrix} \frac{l_1}{\sqrt{l_1^2+b^2l_2^2}} & -\frac{bl_2}{\sqrt{l_1^2+b^2l_2^2}} \\ \frac{bl_2}{\sqrt{l_1^2+b^2l_2^2}} & \frac{l_1}{\sqrt{l_1^2+b^2l_2^2}} \end{pmatrix}$$

Now suppose that

$$B_1 = \begin{pmatrix} \frac{1}{l_1} & 0 \\ 0 & \frac{1}{l_2} \end{pmatrix}$$

$$B_2 = \begin{pmatrix} l_{f_1} & 0 \\ 0 & l_{f_2} \end{pmatrix}$$

and

$$B_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Then

$$B_1R_1B_2B_3 = B$$

Thus the linear transformation at the heart of SCA is equivalent to rescaling d_1 and d_2, orthogonally rotating by an angle $\theta + \alpha$, rescaling f_1 and f_2, and finally reflecting d_1 and d_2. The form of R_1 is the same as that of R. Thus the SCA method is similar to the Jacobi method; the difference is in the angle and the rescalings. That is why the SCA method is called a Jacobi-like method.

3.7 Some useful results about the difference (α) between the SCA angle and the Jacobi angle (θ)

In this section, the relationship of α (the difference between the SCA angle and the Jacobi angle) to θ (the Jacobi angle) is going to be discussed. Compare R_1 with R_J (where R_J is the 2 by 2 rotation matrix for the Jacobi method). It is known in Section 3.6 that $|(\theta + \alpha)| < \pi/2$. The

difference, α, between the angles of the SCA method and that of the Jacobi method is as follows.

$$\begin{aligned}
\sin 2\alpha &= \sin[2(\theta+\alpha)-2\theta] \\
&= \sin 2(\theta+\alpha)\cos 2\theta - \cos 2(\theta+\alpha)\sin 2\theta \\
&= 2\sin(\theta+\alpha)\cos(\theta+\alpha)\cos 2\theta - (\cos^2(\theta+\alpha)-\sin^2(\theta+\alpha))\sin 2\theta \\
&= \left(\frac{1}{l_1^2+b^2l_2^2}\right)\left(2bl_1l_2\cos 2\theta - (l_1^2-b^2l_2^2)\sin 2\theta\right). \qquad (3.14)
\end{aligned}$$

So from equation (3.14) the difference between the SCA angle and the Jacobi angle depends on the length of d_1 and d_2, the parameter b and the Jacobi angle θ. The difference α between the SCA angle and the Jacobi angle is important, but the sign is not.

Now I prove that the SCA direction is the one with minimal $|\alpha|$ within the set of possible b. Recall equation (3.8) in Section 3.2,

$$v_{11}^{(1)} = \left(\frac{v_{11}-v_{22}}{2}\right)\cos 2(\theta+\alpha) + v_{12}\sin 2(\theta+\alpha) + \frac{(v_{11}+v_{22})}{2}. \qquad (3.15)$$

The term $\frac{(v_{11}+v_{22})}{2}$ is constant, so we just consider the first two parts of equation (3.15). In this case, θ is constant, but the first two parts can be thought of as a function $g(\alpha)$.

$$\begin{aligned}
g(\alpha) &= \frac{(v_{11}-v_{22})}{2}\cos 2(\theta+\alpha) + v_{12}\sin 2(\theta+\alpha) \\
&= \frac{(v_{11}-v_{22})}{2}(\cos 2\theta\cos 2\alpha - \sin 2\theta\sin 2\alpha) + v_{12}(\sin 2\theta\cos 2\alpha + \cos 2\theta\sin 2\alpha) \\
&= \left(\frac{(v_{11}-v_{22})}{2}\cos 2\theta + v_{12}\sin 2\theta\right)\cos 2\alpha - \left(\frac{(v_{11}-v_{22})}{2}\sin 2\theta - v_{12}\cos 2\theta\right)\sin 2\alpha.
\end{aligned}$$

The Jacobi angle θ lies in the range $|\theta| \leq \pi/4$. When $v_{11} > v_{22}$, $\tan 2\theta = \frac{2v_{12}}{(v_{11}-v_{22})}$.

According to above the equation,

$$\begin{aligned}
\cos 2\theta &= \frac{1}{\sqrt{1+\tan^2 2\theta}} = \frac{(v_{11}-v_{22})}{\sqrt{(v_{11}-v_{22})^2+4v_{12}^2}} \\
\text{and} \quad \sin 2\theta &= \tan 2\theta\cos 2\theta = \frac{2v_{12}}{\sqrt{(v_{11}-v_{22})^2+4v_{12}^2}}.
\end{aligned}$$

So that

$$g(\alpha) = \left(\frac{(v_{11}-v_{22})^2}{2\sqrt{(v_{11}-v_{22})^2+4v_{12}^2}} + \frac{2v_{12}^2}{\sqrt{(v_{11}-v_{22})^2+4v_{12}^2}}\right)\cos 2\alpha$$

$$+\left(\frac{-v_{12}(v_{11}-v_{22})}{\sqrt{(v_{11}-v_{22})^2+4v_{12}^2}}+\frac{v_{12}(v_{11}-v_{22})}{\sqrt{(v_{11}-v_{22})^2+4v_{12}^2}}\right)\sin 2\alpha$$

$$= \cos 2\alpha\frac{\sqrt{(v_{11}-v_{22})^2+4v_{12}^2}}{2}.$$

The term $\frac{\sqrt{(v_{11}-v_{22})^2+4v_{12}^2}}{2}$ is a constant and $\cos 2\alpha$ is a even function. So, for any α, $g(\alpha) = g(-\alpha)$. Also for any k, $|\alpha| \leq \pi/4$. In this range, when $\alpha = 0$, $g(\alpha)$ attains its maximum value and as $|\alpha|$ increases, $g(\alpha)$ decreases. All the above process and equation (3.13) can be extended to SCA and Jacobi angles at any iteration. So the best choice of b at any iteration i is that one which minimizes $|\alpha|$. For any given directions r and s, the SCA angle is the one with minimal $|\alpha|$. So in every iteration, the SCA angle is the one with minimal $|\alpha|$. However what has not been shown is that out of all pairs of directions to transform, the best pair to choose is the one for which $|\alpha|$ is minimized.

3.7.1 The relationship between $|\alpha|$ and θ

Now consider the relationship between $|\alpha|$ and θ according to equation (3.14), where α is the difference between the SCA angle and the Jacobi angle and θ is the Jacobi angle.

Figure 3.2 shows that the number of zero points ($\alpha = 0$) for $l_1/l_2 = 1, 3, 5$ and $k = 0$ is 3 but the positions (θ) of the zero points are different for different l_1/l_2. The number of the zero points is 3 because when $k = 0$ the possible values of b are 0 and ± 1 and each b value corresponds to one zero point. Angle 0 is the same zero points for $l_1/l_2 = 1, 3, 5$ because when $b = 0$, $\theta = 0$, α must be 0 for any l_1 and l_2 (equation (3.14)). Furthermore, according to equations (3.12) and (3.13), $\tan(\theta + \alpha) = bl_2/l_1$. So for the zero points $\tan\theta = bl_2/l_1$, for the same b (b is not 0). If l_1/l_2 is different, the positions (θ) are different because the angle θ depends on b and l_2/l_1. That is why the position of the zero point is different for different l_1/l_2 when b is not zero. The higher the ratio l_1/l_2, the nearer are the other two 0 points to angle 0°.

Figure 3.3 shows that the higher the k, the more zero points. This can be expected because as k increases, the number of directions available for SCA algorithm increases. The same zero points −45°, 0°, and 45° occur for different k, because these directions are the same for different k and in these positions $\alpha = 0$ for different k. The higher the k, the smaller the $|\alpha|$. This can be

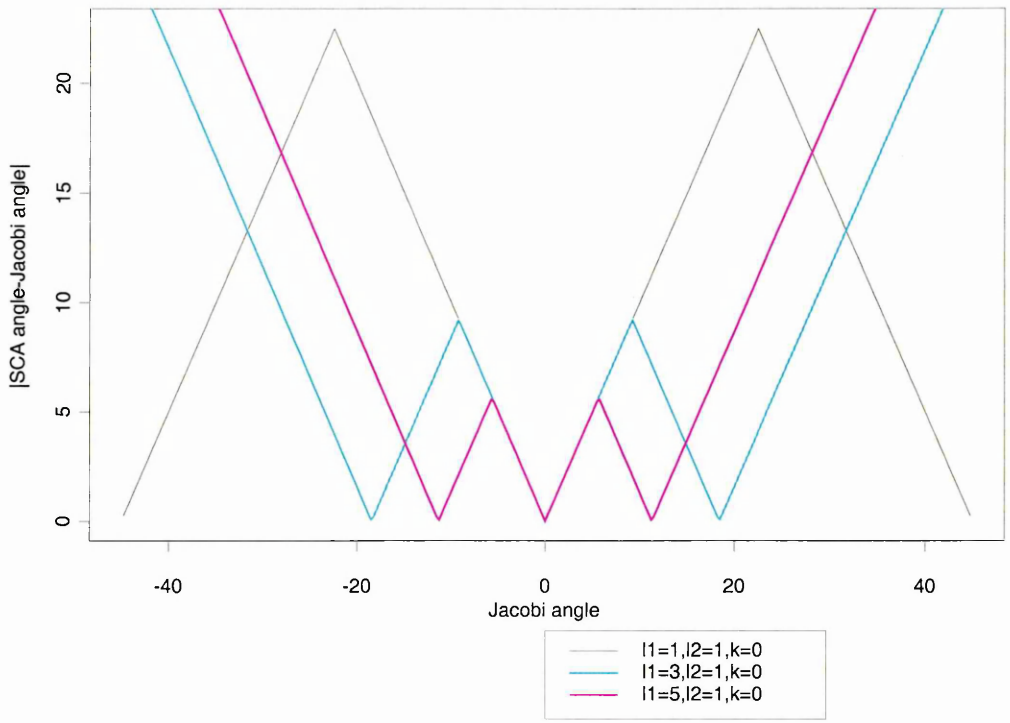

Figure 3.2: The relationship of $|\alpha|$ and θ when $\frac{l_1}{l_2}$=1, 3 and 5

expected because as k increases, there are more SCA directions and the neighbouring two SCA directions become nearer as demonstrated by Figure 3.1. All in all, as k increases, there are more 0 points, the absolute value of, $|\alpha|$, the difference between the angle of the SCA method and that of the Jacobi method, is smaller.

3.8 Discussion and conclusion

In this chapter, an iterative algorithm has been introduced. This algorithm produces simple components, which can be considered as approximations to the principal components. In the construction of simple components, the variance each component explains, the size of angles between the simple components and corresponding PCs and the easier interpretation of components are considered. So the simple components algorithm is different from most of the algorithms

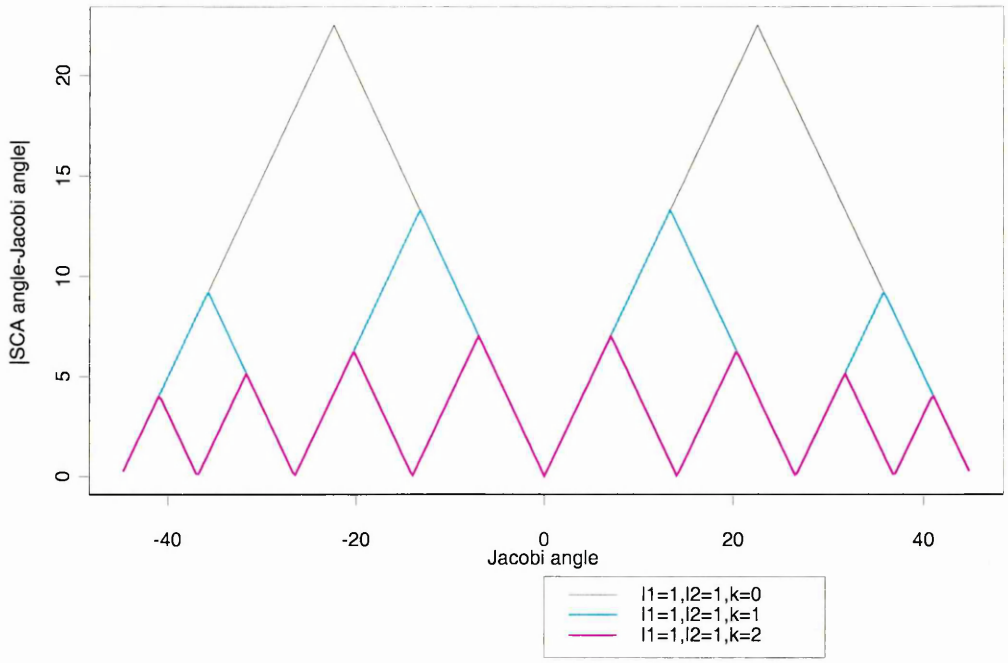

Figure 3.3: The relationship of $|\alpha|$ and θ when k is 0, 1 and 2

in Chapter 2 because most of the algorithms in Chapter 2 just consider the variance each component explains and the easier interpretation of components but the size of the angles between the modified components and original PCs is unimportant. For example, the approximation of the principal components by the simple components produced by SCA methods for the pitprop data is not so good as that for the RI data and the Sparrow data because the angles for the pitprop data are too big compared to the angles for the RI data and Sparrow data. However the angles obtained by the SCA algorithm for this data are less than the angles obtained by SCoTLASS, SCA_{RG} and SPCA given in Chapter 2. This is because SCoTLASS, SCA_{RG} and SPCA ignore the size of the angles between the principal components and the corresponding modified components. From all the examples in Chapter 3, it was shown that simple components can be very good approximations to the PCs for some data sets. For example, for the RI data, the angles between simple components and their corresponding principal components are very

small. Also each simple component accounts for a similar amount of variance as the corresponding PCs. Furthermore, the simple components produced are generally easier to interpret than the equivalent PCs because of the small valued integer loadings.

The performance of the SCA algorithm is controlled by the nonnegative integer k, which determines the number of directions considered for each simplicity preserving transformation at each iteration. As k increases, the directions available to get simple components increases. For the examples in Chapter 3 $k = 0$ is a good choice.

Six different variations of the SCA algorithm such as multiple updating with maximal variance or maximal improvement in variance or single updating with maximal variance or maximal improvement in variance were introduced in this chapter. From the results obtained from different SCA methods, it was found that different simple components can be obtained for the same dataset. For example, for the pitprop data in Section 3.5.4, the simple components obtained by SCA2 (single SCA method with maximal improvement in variance) are close to the corresponding PCs and much easier to interpret than the simple components obtained by SCA5 (multiple SCA methods with maximal improvement in variance). So these six different SCA methods will be investigated in the following chapters to see which SCA methods are good.

Chapter 4

Population results for 6 dimensional data - $k = 0$

4.1 Introduction

This chapter is concerned with the results produced by SCA methods applied to the population variance-covariance matrix V_0 for 6 different 6 dimensional data sets (the definition of V_0 is given in next section), the six different SCA methods are introduced in Section 3.6. In this chapter, 6 dimensional data sets will be investigated because the dimension 6 is not too big and not too small, this kind of data should reveal the essence of the SCA methods; Jolliffe and Uddin (2002); considered 6 dimensional data sets. It is necessary for a good SCA method to perform well on the population covariance matrix because PCA retrieves the original eigenvectors' matrix and a good SCA method should get a good approximation for PCA. The results of SCA methods applied to V_0 are called population results.

Section 4.2 introduces the structure of the data used in this chapter. Section 4.3 gives the specific data used. Section 4.4 introduces how to measure the accuracy and simplicity of the results obtained by the SCA methods. Considering that generally $k = 0$ seems to be a good choice for SCA methods to get good results (Section 3.8), Section 4.5 will provide the population results when $k = 0$ for the simple structures. Section 4.6 will provide the population results

for the complex structures when $k = 0$. The last section, Section 4.7, gives a discussion and conclusion.

4.2 Structures of the data

Let $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_p)$ be a vector which consists of ordered real positive numbers and A be an orthogonal matrix, i.e. $A^T A = I$, where I is a $p \times p$ unity matrix. Then a variance covariance matrix whose eigenvalues are λ_i $(i = 1, 2, \ldots, p)$ with corresponding eigenvectors a_i $(i = 1, 2, \ldots, p)$, the i^{th} column of matrix A, can be constructed via the formula $V_0 = \sum_{i=1}^{p} \lambda_i a_i a_i^T$.

In this Chapter, 3 types of structures of the eigenvectors will be investigated. Following Jolliffe and Uddin (2002), these 3 structures are block structure, uniform structure and intermediate structure. So when investigating these structures, hopefully before using SCA methods, just from the structures of PCs, it will show whether the SCA methods work well or not for each structure, because generally SCA methods are used, when PCA can not give satisfactory interpretation of principal components.

4.2.1 Block structure

In block structures, the elements of the eigenvector in the block structure are either close to zero or far from zero. So, after the columns and rows have been rearranged, the structure takes the approximate form

$$A = \begin{pmatrix} A_1 & 0 & \ldots & 0 \\ 0 & A_2 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & A_m \end{pmatrix}.$$

where all the 0 matrices in column k have the same number of column as A_k $(k = 1, 2, \ldots, m)$ and all the 0 matrices in row k have the same number of rows as A_k $(k = 1, 2, \ldots, m)$. Suppose i_k and j_k are the numbers of rows and columns of matrix A_k respectively. Then $i_1 + i_2 + \ldots + i_m = p$

and $j_1 + j_2 + \ldots + j_m = p$ where p is the dimension of the structure. The smaller the i_k and j_k, the easier the interpretation for this structure.

4.2.2 Uniform structure

If the absolute values of elements in each column of the eigenvector matrix A are very similar, the eigenvector matrix said to have a uniform structure. It is hard to make all the columns in a matrix orthogonal and the absolute values of all the elements in each column very similar simultaneously. So in our example, only the first two columns of the uniform structure have the features of this structure.

4.2.3 Intermediate structure

Intermediate structures are taken to be structures between block structure and uniform structure, where all the elements in the components are intermediate in value and vary in a wide range if the columns of eigenvector matrix are normalized. Generally, only the first two columns of the intermediate structure have the features of this structure.

4.2.4 Simple structure and complex structure

If all the elements in the structure can be represented by small-valued integers (the mean of the absolute value of all the integer elements in each column of the eigenvector matrix is less than 10), the structure is called a simple structure. Complex structures cannot be represented by small valued integers. However, each column of a complex structure may still be required to be close to (have similar directions) the corresponding column of the simple structure, i.e. the angle between each column of simple structure and the corresponding column of the complex structure is generally smaller than the angles between each column of simple structure and each column of all the other complex structures. In other words, by definition complex structure is the closest complex structure to the corresponding simple structure.

4.3 Specific data used

In the previous section, the definitions of the three structures were given. The results obtained by SCA methods are expected to enhance the interpretation of PCs generally. The loadings vectors produced by the SCA algorithm must be integers. So if a SCA method is good for some structure, it should retrieve a good approximation to the simple form of this structure. For this reason, first the six SCA methods are applied to the variance covariance matrices for the simple structures. The same vector of eigenvalues for all the structures, λ, $(12, 8, 6, 4, 2, 1)$ are used. So all the eigenvalues are different, and the corresponding eigenvectors are uniquely defined (Jolliffe (2002b) Sections 2.1, 2.2, 3.1 and 3.2) and the first two components are important because they explain about 61% of variance.

Block structure, uniform structure and intermediate structure will be investigated in this chapter. Each structure will have both a simple form and complex form. In the following examples, the normalizing constant for the simple structures is omitted.

The first example is simple block structure:

$$A = \begin{pmatrix} 0 & -1 & 1 & 0 & 1 & 0 \\ 0 & -1 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & -2 & 0 \\ 1 & 0 & 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & -1 & 0 & -1 \\ 1 & 0 & 0 & 2 & 0 & 0 \end{pmatrix}.$$

After exchanging the first column with the fifth column, and taking a 3 by 3 matrix are as an element, the matrix is block-diagonal. As all the elements are integers, A has a simple block structure.

The second example is a simple uniform structure:

$$A = \begin{pmatrix} -1 & 1 & 0 & 1 & -1 & 0 \\ -1 & 1 & 0 & -1 & -1 & 0 \\ -1 & 1 & 0 & 0 & 2 & 0 \\ 1 & 1 & 0 & 0 & 0 & 2 \\ 1 & 1 & -1 & 0 & 0 & -1 \\ 1 & 1 & 1 & 0 & 0 & -1 \end{pmatrix}.$$

The absolute values of the first two components are the same, so it's a simple uniform structure.

The third example is simple intermediate structure:

$$A = \begin{pmatrix} 1 & -2 & 2 & 1 & -1 & 1 \\ 1 & -2 & -1 & -2 & -1 & 0 \\ 1 & -2 & -1 & 1 & 2 & -1 \\ 2 & 1 & -1 & 0 & 1 & 2 \\ 2 & 1 & -1 & 1 & -2 & -1 \\ 2 & 1 & 2 & -1 & 1 & -1 \end{pmatrix}.$$

All the elements here are integers and intermediate in value when the columns of A are normalized, so it's a simple intermediate structure.

For this study, the complex structures are close to the corresponding simple structures (i.e. the directions are similar). The complex block and intermediate data come from Jolliffe and Uddin (2002). We do not use the complex uniform structure given by Jolliffe and Uddin (2002) because the following complex uniform structure is closer to the simple uniform structure of this section.

The fourth example is complex block structure:

$$A = \begin{pmatrix} 0.10 & -0.54 & 0.76 & -0.12 & 0.34 & -0.02 \\ 0.08 & -0.57 & -0.60 & 0.23 & 0.51 & -0.01 \\ 0.08 & -0.61 & -0.12 & -0.12 & -0.77 & 0.02 \\ 0.59 & 0.09 & -0.07 & -0.31 & 0.07 & 0.73 \\ 0.58 & 0.10 & -0.11 & -0.42 & 0.05 & -0.68 \\ 0.53 & 0.07 & 0.18 & 0.81 & -0.16 & -0.07 \end{pmatrix}.$$

If the small values (absolute values less than 0.25) are taken as zeroes, and the first column is exchanged with the fifth column, it is clear that this is a complex block structure. The angles between the simple block structure and complex block structure are 9, 9, 16, 17, 13 and 5 degrees respectively.

The fifth structure is the complex uniform structure

$$A = \begin{pmatrix} -0.42 & 0.38 & 0.07 & 0.74 & -0.33 & 0.13 \\ -0.40 & 0.41 & 0.01 & -0.63 & -0.49 & -0.18 \\ -0.37 & 0.46 & 0.03 & -0.11 & 0.79 & 0.10 \\ 0.48 & 0.41 & 0.11 & -0.14 & -0.13 & 0.74 \\ 0.34 & 0.40 & -0.78 & 0.11 & 0.01 & -0.31 \\ 0.43 & 0.38 & 0.61 & 0.10 & 0.04 & -0.55 \end{pmatrix}.$$

The absolute values of the elements in the first two components are very similar. The first two components approximate $(-1,-1,-1,1,1,1)^T$ and $(1,1,1,1,1,1)^T$ respectively. However the first two components can not be represented by integers, so it is a complex structure. The angles between the simple uniform structure and complex uniform structure are 6, 4, 11, 14, 10 and 17 degrees respectively.

The last structure is complex intermediate structure

$$A = \begin{pmatrix} 0.22 & -0.51 & 0.60 & 0.30 & -0.33 & 0.36 \\ 0.25 & -0.52 & -0.36 & -0.64 & -0.34 & -0.06 \\ 0.23 & -0.55 & -0.25 & 0.38 & 0.61 & -0.27 \\ 0.55 & 0.25 & -0.25 & -0.05 & 0.26 & 0.71 \\ 0.52 & 0.25 & -0.26 & 0.45 & -0.51 & -0.37 \\ 0.51 & 0.20 & 0.56 & -0.38 & 0.28 & -0.40 \end{pmatrix}.$$

The elements are intermediate in value, so it is complex intermediate structure. The angles between the simple intermediate structure and complex intermediate structure are 3, 4, 6, 8, 6 and 7 degrees respectively.

4.4 Measurement of accuracy and simplicity

The data structures in Section 4.3 will be used to investigate if the SCA methods in Section 3.4 are good in returning these structures, and to assess the different methods to find which one is the most accurate, which one is the most simple. Two criteria are used to measure the performance of each SCA method. One criterion is based on angles. This is the absolute value of the angles between the original components and the corresponding simple components. So the smaller the angles, the more accurate the results. The other criterion is the mean of the absolute value of all the elements in a component. So the smaller the mean, the more simple the results. If the mean of the component is less than 10, in general, the component is easy to interpret. So the component is simple, otherwise it's complex. However, in practice the difference between a simple solution and a complex solution is not very clear, a component with high mean might be easier to interpret than a component with small mean. This criterion only can compare the simplicity of different simple components with each other, but not with competitors.

4.5 Results on population covariance matrix V_0-simple structures

First all the SCA methods are applied to the variance covariance matrices V_0 for the simple structures. In this chapter, the results of PCA are not displayed because by definition PCA must retrieve V_0. Only the first two components of the following examples have the properties of the structure and they account for about 61% of variance, so only the first two components are discussed. A result is good or bad for a structure if it is good or bad for the first two components of the structure.

4.5.1 Simple block structure

Now I am going to investigate the population results obtained by SCA methods for simple block structure. Three of the SCA methods get the simple block structure back exactly: SCA2, SCA5

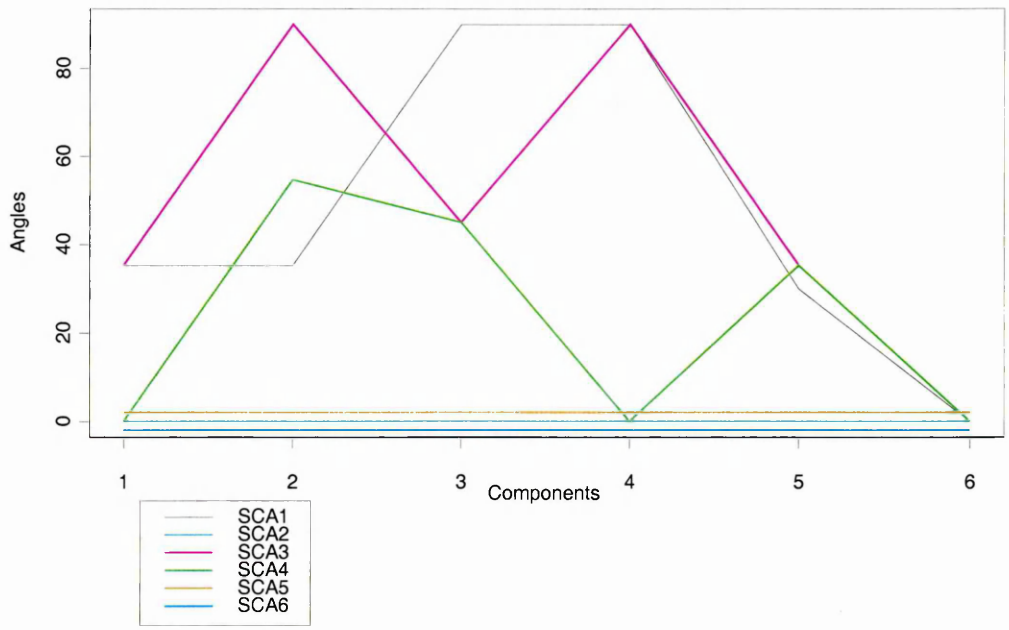

Figure 4.1: The accuracy of components produced by different SCA methods for 6 dimensional simple block structure

and SCA6 (Figure 4.1). So they are the best methods for this example of simple block structure. In order to distinguish them, I add 2° for each angle for SCA5 and subtract 2° for each angle for SCA6 in figure 4.1. SCA4 retrieve the first component. The results of SCA1 are disappointing and the results of SCA3 are even more disappointing for the first two components because the angles here are too big. Checking the results of SCA methods, all the SCA methods preserve the block structure.

Why did SCA2, SCA5 and SCA6 get the same results? SCA2 took five iterations, SCA5 and SCA6 took three iterations.

In the five iterations, SCA2 updates columns 4 and 5, 4 and 6, 1 and 3, 1 and 2, 2 and 3 respectively. The corresponding transformation matrices can be represented by P_1, P_2, ..., P_5. So the transformations performed by SCA2 are $P_1P_2P_3P_4P_5$. The pairs of columns updated by

SCA2 tie in with the simple block structure. For example, the transformation matrix performed by SCA2 in the first step,

$$P_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

In the first iteration, SCA5 and SCA6 both update columns 4 and 5 and columns 1 and 3, and the transformation matrix of SCA5 and SCA6 in the first iteration is equivalent to P_1P_3. In the second iteration, SCA5 and SCA6 update columns 1 and 2 and columns 4 and 6. The corresponding transformation matrix is equivalent to P_2P_4. In the third iteration, SCA5 and SCA6 update columns 2 and 3, the corresponding transformation matrix is P_5. Now SCA2 updates different pairs of columns in iterations 2 and 3, so P_2 and P_3 can be exchanged. So SCA5 and SCA6 gave the same answer as that of SCA2 even though pairs of columns are updated in a different order.

Comparing the accuracy of the results obtained by SCA2 with SCA4, and SCA5 with SCA6, Figure 4.1 shows that the results using maximal improvement in variance are at least as accurate as those using maximal variance. The results using the maximal improvement criterion and single updating (SCA2) are the same as those obtained using multiple updating (SCA5). The results using the maximal variance and multiple updating (SCA6) are more accurate than those with single updating (SCA4). The condition $l_r = l_s$ makes the results obtained by SCA1 and SCA3 less accurate than those obtained by corresponding SCA methods (the results obtained by SCA1 are less accurate than those obtained by SCA2, the results obtained by SCA3 are less accurate than those obtained by SCA4).

All the means of the components obtained by SCA methods for the simple block structure are much smaller than 10 (Figure 4.2), so the components produced by all the SCA methods are very simple. SCA3 is the simplest method. The means of the components obtained by SCA1, SCA3 and SCA4 are smaller than the actual means, so they are too simple. The means

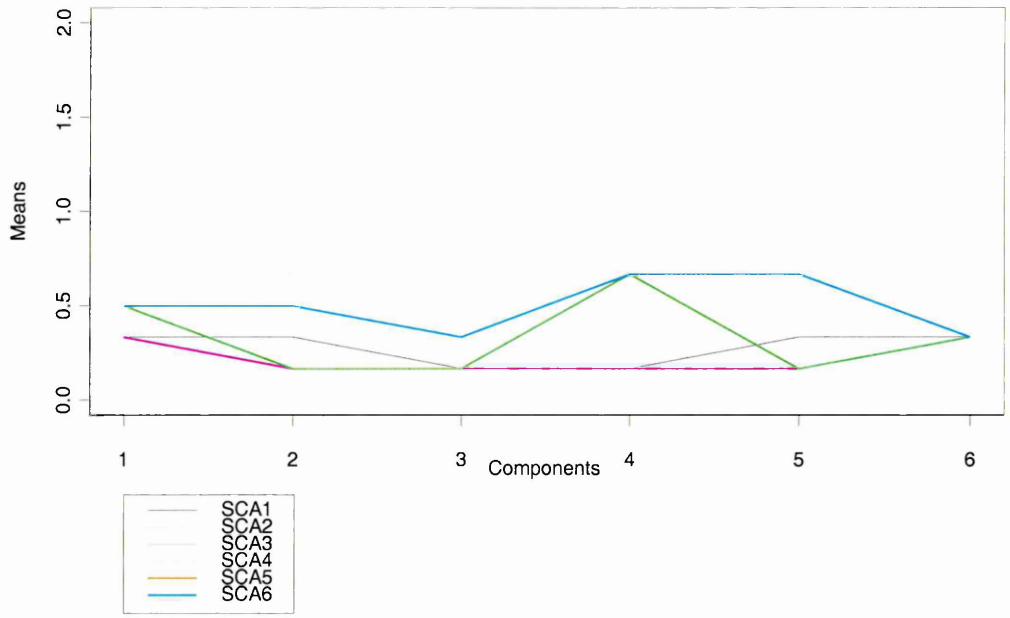

Figure 4.2: The simplicity of components produced by different SCA methods for 6 dimensional simple block structure

of the components obtained by SCA2, SCA5 and SCA6 are the same as the actual means. This is expected because SCA2, SCA5 and SCA6 get the simple block structure back exactly. The components produced by SCA methods with maximal variance are at least as simple as those produced by SCA methods with maximal improvements in variance (the components produced by SCA3 and SCA4 are simpler than those produced by SCA2, the components produced by SCA6 are the same as those produced by SCA5). The components produced by SCA methods of single updating are at least as simple as those obtained by multiple updating (the components produced by SCA3 and SCA4 are simpler than those produced by SCA6, the components produced by SCA1 are simpler than those produced by SCA5, the components produced by SCA2 are the same as those produced by SCA5). This could be explained by the number of steps each method took because the transformed directions are more complex than the directions in the

previous iteration (Sections 3.3.2, 3.3.3). SCA2 took five iterations (five steps). SCA5 took 3 iterations (five steps, the first two iterations have two steps, the third has one step). SCA6 took 3 iterations (five steps). SCA1, SCA3 and SCA4 took two iterations (two steps). SCA1, SCA3 and SCA4 seems to stop too early, this stopped them reaching the same answer as SCA2, SCA5 and SCA6. SCA2, SCA5 and SCA6 took the same number of steps. The more steps a SCA method took, generally the more complex the results.

Finally, combining the simplicity and accuracy of the results obtained by SCA methods, SCA2, SCA5 and SCA6 are the best methods for this structure.

4.5.2 Simple uniform structure

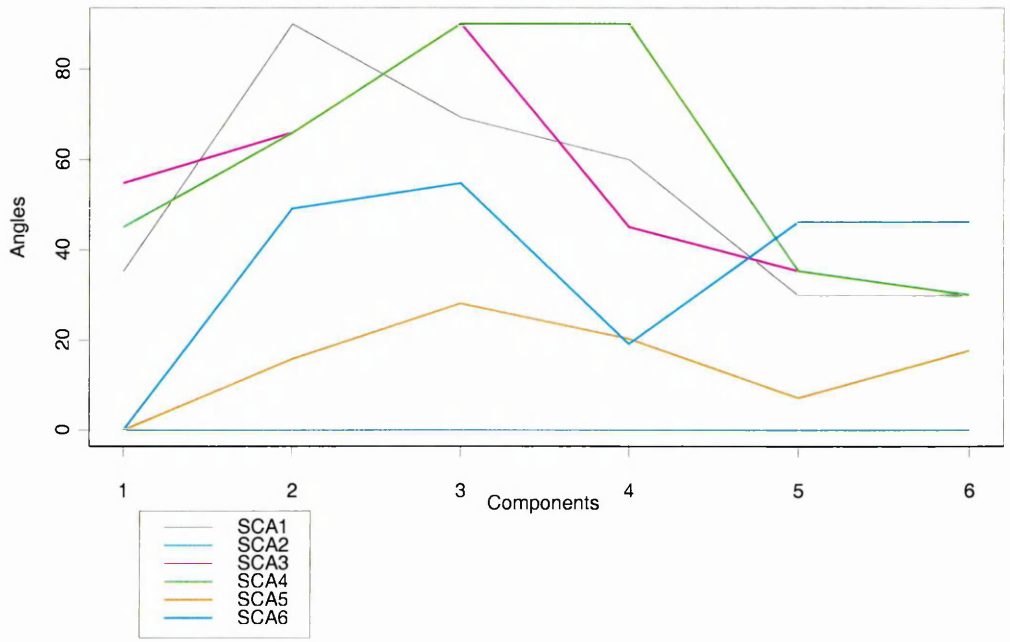

Figure 4.3: The accuracy of components produced by different SCA methods for 6 dimensional simple uniform structure

Considering the population results for simple uniform structure. Only one method, SCA2

(maximal improvement in variance with single updating) gave the correct results (Figure 4.3). So SCA2 is the most accurate method for simple uniform structure. The second most accurate method is SCA5, the third most accurate method is SCA6. SCA3 got the least accurate results for the first component. SCA1 got the least accurate results for the second component. Checking the actual components obtained by SCA methods, only SCA2 preserves the uniform structure.

Comparing the accuracy of the results obtained by SCA2 with SCA4, SCA5 with SCA6, Figure 4.3 shows that the results using maximal improvement in variance are more accurate than those using maximal variance. The results of single updating with maximal improvement in variance are more accurate than those of multiple updating with maximal improvement in variance (the results obtained by SCA2 are more accurate than those obtained by SCA5). However, the results of multiple updating with maximal variance are more accurate than those of the single updating with maximal variance (the results obtained by SCA6 are more accurate than those obtained by SCA4).

Then, considering the simplicity of the population results produced by SCA methods, the means of the first two components obtained by all the SCA methods are much smaller than 10, so all the SCA methods are simple for the first two components (Figure 4.4). SCA2 gets the correct answer, and the components produced by SCA2 are more simple than those produced by SCA5 and SCA6. The components produced by SCA1, SCA3 and SCA4 are more simple than the correct answer. SCA3 is the simplest method. The components produced by SCA methods with maximum variance are more simple than those produced by SCA methods with maximal improvement in variance (the components produced by SCA4 are more simple than those produced by SCA2, the components produced by SCA6 are more simple than those produced by SCA5). The components produced by single updating SCA methods are more simple than those obtained by multiple updating SCA methods (the components obtained by SCA1, SCA2, SCA3, SCA4 are more simple than those obtained by SCA5, SCA6). This is expected considering the number of steps each method took. SCA1 took four iterations (4 steps). SCA2 took seven iterations (7 steps). SCA3 took just one iteration (1 step). SCA4 took two iterations (2 steps). SCA5 took eight iterations (17 steps). SCA6 took six iterations (8 steps). So the multiple methods generally took more steps. Condition $l_r = l_s$ makes the calculation of SCA1

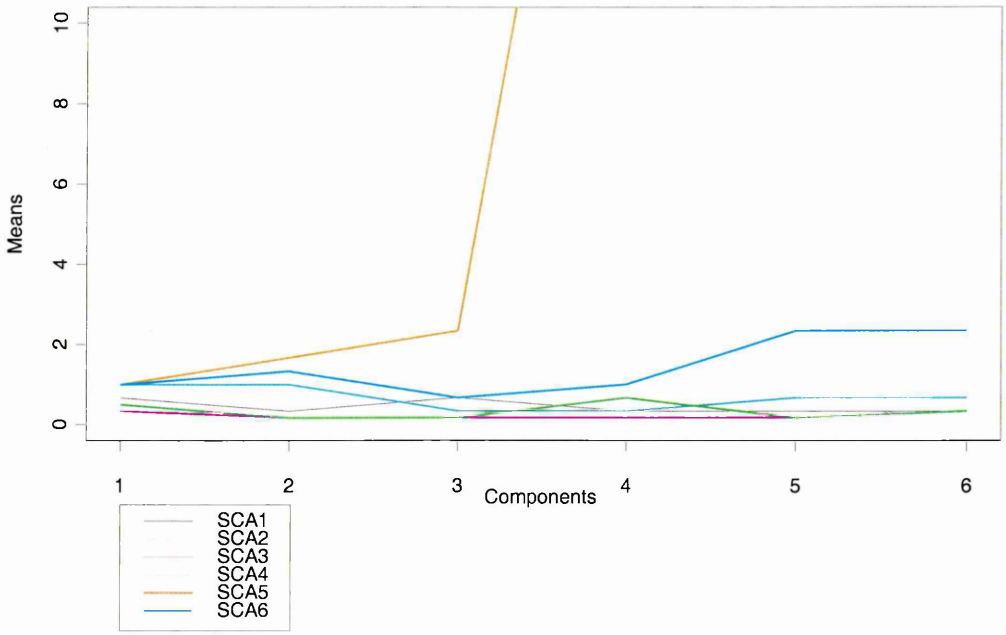

Figure 4.4: The simplicity of components produced by different SCA methods for 6 dimensional simple uniform structure

and SCA3 stop earier than that of the corresponding SCA methods without this condition. This reason makes the results obtained by SCA1 and SCA3 less accurate but more simple than those obtained by other SCA methods. The number of steps taken by SCA2 is more than that taken by SCA1 and the number of steps taken by SCA4 is more than that taken by SCA3. The number of steps taken by SCA5 is more than that taken by SCA2 and the number of steps taken by SCA6 is more than that taken by SCA4. The number of steps taken by SCA5 is more than that taken by SCA6. So as I said in Section 4.5.1, the more steps each method took generally implyes more complex final results.

Finally, combining the simplicity and accurate of the results obtained by SCA methods, SCA2 is the best method for this structure.

4.5.3 Simple intermediate structure

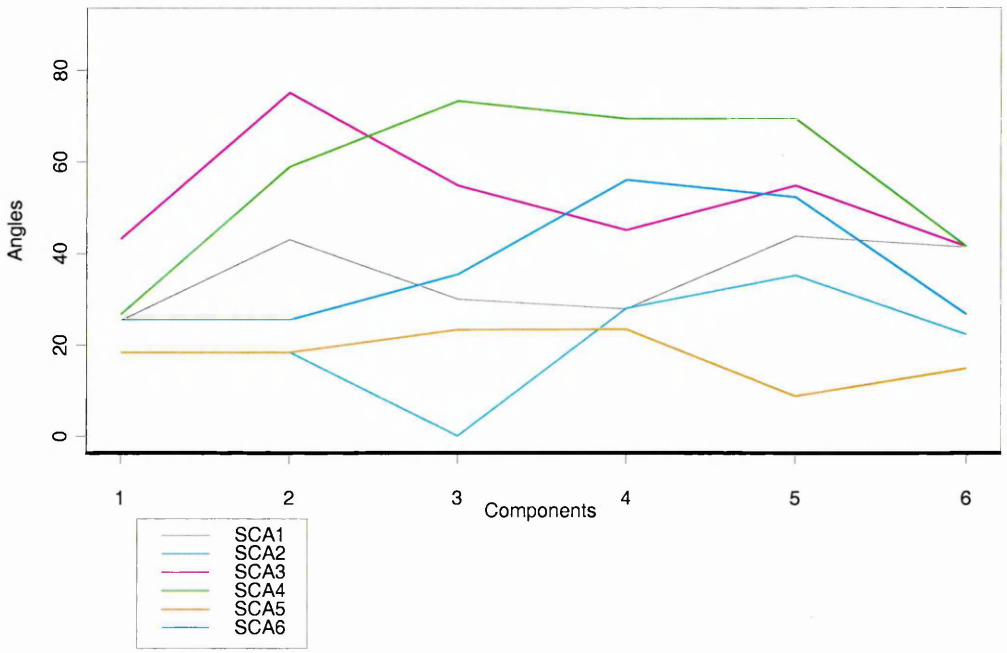

Figure 4.5: The accuracy of components produced by different SCA methods for 6 dimensional simple intermediate structure

In this subsection, I am going to investigate the population results for simple intermediate structure. Interestingly, when $k = 0$, none of the SCA methods were able to recover the simple intermediate structure using the exact covariance matrix. One explanation is that given the set of pairwise transformations available to the algorithm, it is simply not possible for the simple intermediate structure to be obtained. Nevertheless the method SCA2 performs the best out of the methods tried (smaller angles between the obtained loadings and the correct loadings, similar variances for the initial components).

As Figure 4.5 shows, for the first two components, SCA2 and SCA5 were the most accurate methods. The accuracy of components produced by other SCA methods is in the order of SCA6, SCA1, SCA4 and SCA3. Comparing the accuracy of the results obtained by SCA2 (maximal

improvement in variance) with that of SCA4 (maximal variance), and SCA5 with SCA6, shows that the results obtained using maximal improvement in variance are more accurate than those obtained using maximal variance. Also the results with maximal improvement in variance of single updating are more accurate than those of multiple updating, but the results obtained by multiple method (SCA6) with maximal variance are more accurate than that obtained by single method (SCA4) with maximal variance.

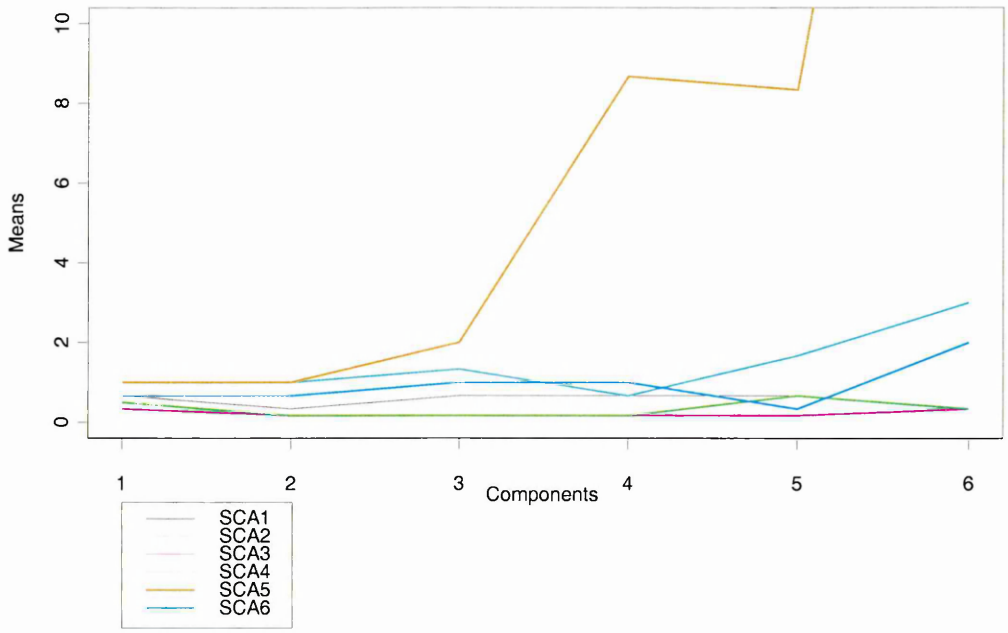

Figure 4.6: The simplicity of components produced by different SCA methods for 6 dimensional simple intermediate structure

Next, check how many iterations each SCA methods took. SCA1 took 5 iterations (5 steps), SCA2 took eight iterations (8 steps). SCA3 took one iteration (1 step). SCA4 took two iterations (2 steps). SCA5 took nine iterations (15 steps). SCA6 took six iterations (7 steps). The number of steps taken by SCA2 is more than that taken by SCA1, the number of steps taken by SCA4 is more than that taken by SCA3. The number of steps taken by SCA5 is more than that taken

by SCA2. The number of steps taken by SCA6 is more than that taken by SCA4. The number of steps taken by SCA5 is more than that taken by SCA6. More steps taken by SCA methods generally mean more complex results.

The first two components produced by the SCA methods are simple for this structure (i.e. the means are less than 10) (Figure 4.6). In fact, the components produced by the SCA methods are more simple than the actual components for the first two components. The results obtained by the SCA methods with maximal variance are more simple than those produced by the SCA methods with maximal improvement in variance. The components obtained by the SCA methods of single updating are more simple than those obtained by multiple updating. This is to be expected judging by the number of steps each method took. In general, the components with fewer steps are more simple.

Finally, combining the simplicity and accuracy of the results obtained by the SCA methods, SCA2 and SCA5 are the best methods for the first two components. Considering that only the first two components have the features of intermediate structure, this means that SCA2 and SCA5 are the best methods for this structure.

4.6 Population results based on population covariance matrix V_0 - complex structures

In this section, the complex structures introduced in Section 4.3 will be investigated. As in Section 4.5, only the first two components are considered.

4.6.1 Complex block structure

For complex block structure, first consider the accuracy of components produced by the SCA methods (Figure 4.7), SCA2 is the most accurate method for the first two components, SCA5 is the second most accurate method, SCA6 and SCA1 are the third most accurate methods for the first two components. SCA4 retrieves the first component, the second component of SCA4 is less accurate. SCA3 is the least accurate method for the first two components. The results

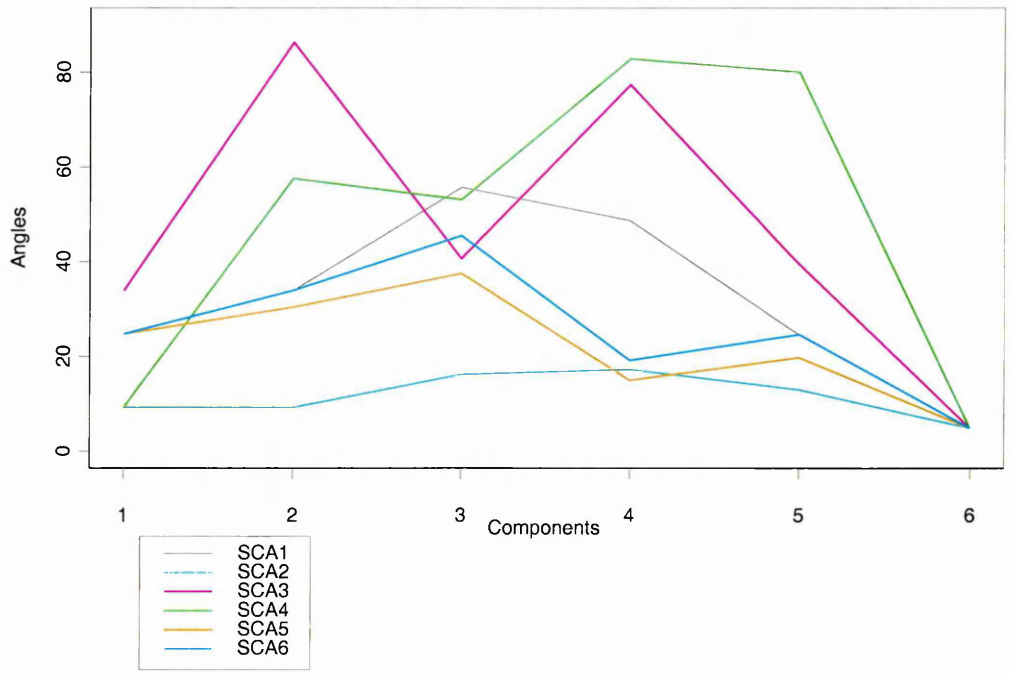

Figure 4.7: The accuracy of components produced by different SCA methods for 6 dimensional complex block structure

obtained by single SCA methods with maximal improvement in variance are more accurate than those obtained by single SCA methods with maximal variance (the results obtained by SCA2 are more accurate than those obtained by SCA4, the results obtained by SCA1 are more accurate than those obtained by SCA3). The results obtained by the multiple method with maximal improvement (SCA5) are more accurate than those obtained by the multiple method with maximal variance (SCA6). The condition $l_r = l_s$ makes the results obtained by SCA1 and SCA3 less accurate than those obtained by corresponding methods without this restriction. The results obtained by the single SCA method with maximal improvement in variance (SCA2) are more accurate than those obtained by the multiple SCA method with maximal improvement in variance (SCA5). These observations are the same as those for the simple block structure. Comparing these results with the results for the simple block structure

(Figure 4.1), the obvious difference is that no methods retrieve the complex block structure. For the complex block structure, the actual components obtained by SCA2 are the same as the original simple block structure (the same components as in simple block structure). So the angles obtained by SCA2 for the first two components are just the angles between the first two components of the simple block structure and the first two components of the complex block structure. It is very surprising that the component 4 obtained by SCA5 is more accurate than component 4 obtained by SCA2. I suppose the simple structures is the closest approximation of the corresponding complex structure. This happened because the fourth component of the simple block structure is not the closest simple component of the corresponding complex structure. The components obtained by SCA3 and SCA4 are the same whether applied to the simple block structure or the complex block structure.

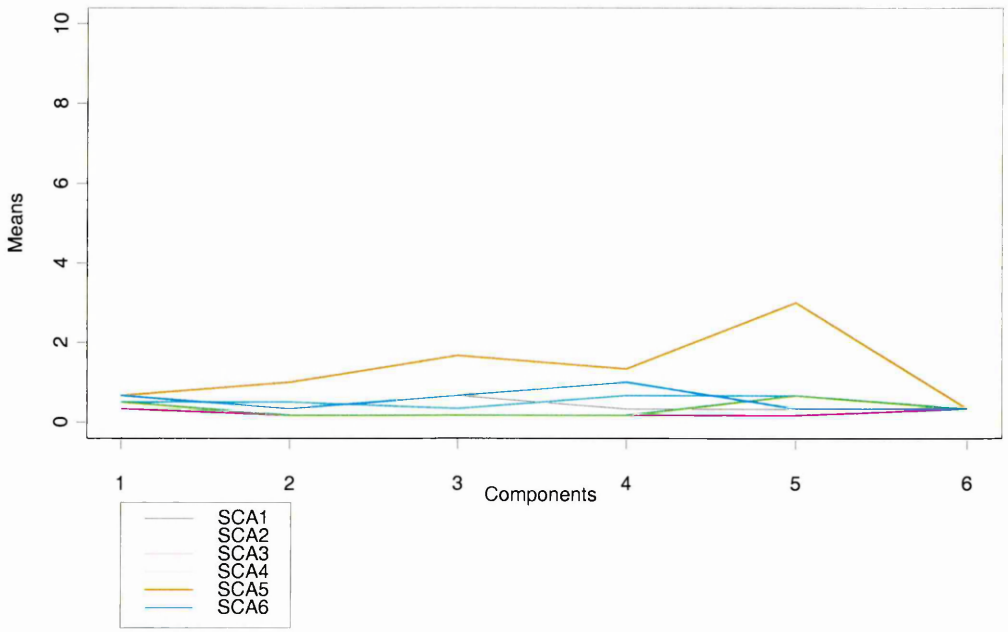

Figure 4.8: The simplicity of components produced by different SCA methods for 6 dimensional complex block structure

The components produced by all the SCA methods are very simple (Figure 4.8). SCA3 and SCA4 are the simplest methods. SCA5 is the least simple method, but the components obtained by SCA5 are still very simple. The components obtained by SCA2 are more simple than those obtained by SCA5. The components obtained by SCA4 are more simple than those obtained by SCA6. The components obtained by SCA4 are more simple than those obtained by SCA2. The components obtained by SCA6 are more simple than those obtained by SCA5. These mean that the components obtained by single SCA methods with maximal improvement in variance are more simple than those obtained by the multiple SCA method with maximal improvement in variance. The components obtained by the single SCA method with maximal variance are more simple than those obtained by the multiple SCA method with maximal variance. The components obtained by the single SCA method with maximal variance are more simple than those obtained by the SCA method with maximal improvement in variance. The components obtained by the multiple SCA method with maximal variance are more simple than those obtained by the multiple SCA method with maximal improvement in variance. These results are the same as those for the simple block structure. This is expected in view of the number of steps each method takes. SCA1 took four iterations (4 steps). SCA2 took five iterations (5 steps). SCA3 took one iteration. SCA4 took two iterations. SCA5 and SCA6 took four iterations (9 steps). Multiple methods took more steps generally. The number of steps taken by SCA2 is more than that taken by SCA1, the number of steps taken by SCA4 is more than that taken by SCA3. SCA5 and SCA6 took the same number of steps.

SCA2 is the best method for this structure given the tradeoff of accuracy and simplicity.

4.6.2 Complex uniform structure

For the first two components of complex uniform structure (Figure 4.9), SCA5 is the most accurate method, the results obtained by SCA2 are similar to the results obtained by SCA5. SCA6 is not very good. SCA4 get the first component back, the second components of SCA4 is less accurate. The results obtained by SCA1 are disappointing, though SCA3 is the least accurate method. The results obtained by SCA methods with maximal improvement in variance are more accurate than those obtained by SCA methods with maximal variance (the results

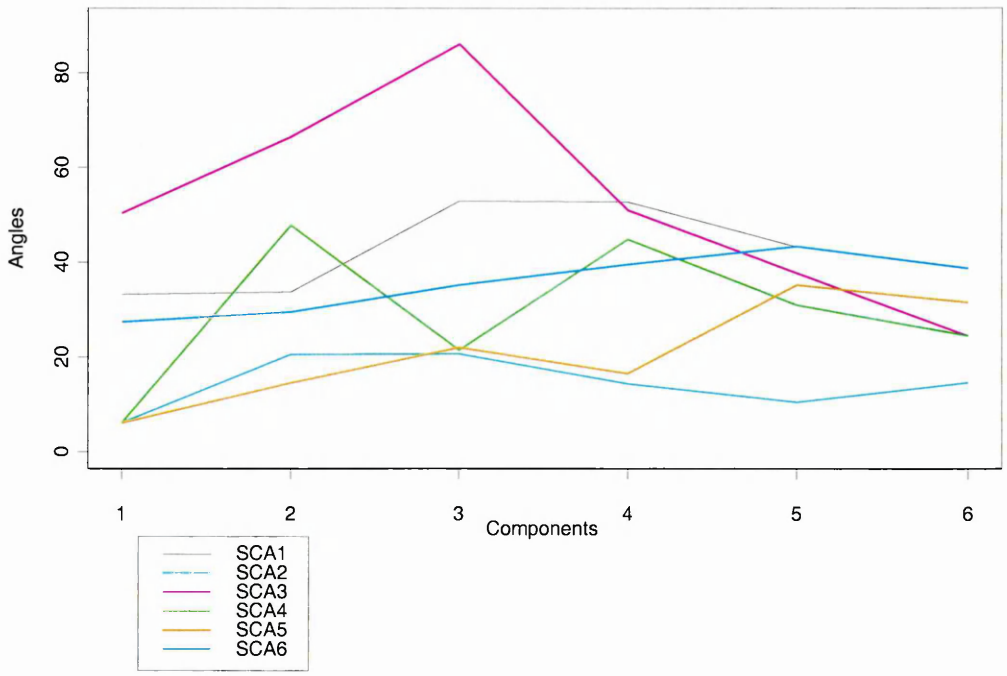

Figure 4.9: The accuracy of components produced by different SCA methods for 6 dimensional complex uniform structure

obtained by SCA2 and SCA5 are more accurate than those obtained by SCA4 and SCA6, the results obtained by SCA1 are more accurate than those obtained by SCA3). The condition $l_r = l_s$ makes the results less accurate. The results obtained by the multiple SCA method with maximal improvement (SCA5) are more accurate than those obtained by the multiple SCA method with maximal variance (SCA6). These observations about the accuracy of the results obtained by SCA methods are the same as those obtained for the simple uniform structure.

SCA1 took five iterations. SCA2 took nine iterations (9 steps). SCA3 took one iteration. SCA4 took four iterations (4 steps). SCA5 took five iterations (12 steps). SCA6 took four iterations (8 steps). So the number of the steps taken by SCA2 is more than that taken by SCA1, the number of the steps taken by SCA4 is more than that taken by SCA3. The number of the steps taken by SCA5 is more than that taken by SCA2. The number of the steps taken

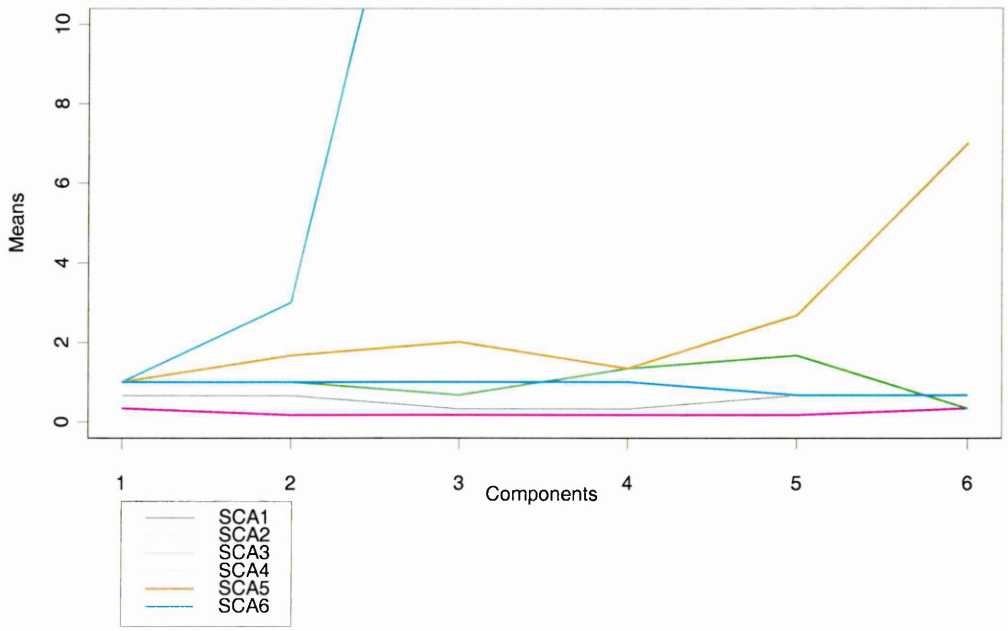

Figure 4.10: The simplicity of components produced by different SCA methods for 6 dimensional complex uniform structure

by SCA6 is more than that taken by SCA4. The number of the steps taken by SCA5 is more than that taken by SCA6. So more steps generally mean less simple results. Considering the simplicity of the components (Figure 4.10), the components obtained by all the SCA methods are very simple for the first two components. All the other observations about simplicity are the same as those for simple uniform structure. This is to be expected in view of the steps taken by each SCA method taken.

So SCA5 and SCA2 are the best methods for this structure given the tradeoff of accuracy and simplicity.

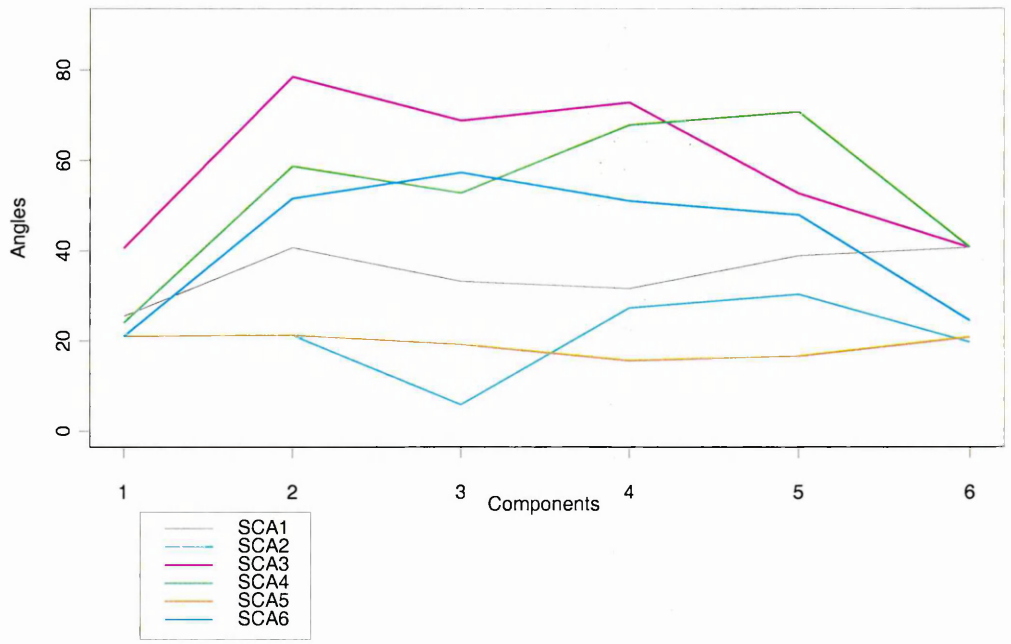

Figure 4.11: The accuracy of components produced by different SCA methods for 6 dimensional complex intermediate structure

4.6.3 Complex intermediate structure

For complex intermediate structure, the results obtained by all the SCA methods are not as good as those obtained by the SCA methods for complex block and complex uniform structures (Figure 4.11). This is not surprising comparing these results with those for simple intermediate structure. For the first two components, SCA2 and SCA5 are the best. The angles obtained by SCA2 and SCA5 for the complex intermediate structure for the first two components are almost the angles between the first two components of simple intermediate structure and the first two components of the complex intermediate structure. So, the results obtained by the SCA methods with maximal improvement in variance are more accurate than those obtained by the SCA methods with maximal variance. The condition $l_r = l_s$ makes the results obtained by SCA1 less accurate than those obtained by SCA2 and the results obtained by SCA3 less accurate

than those obtained by SCA4. The results obtained by the single SCA methods with maximal improvement in variance (SCA2) are the same as those obtained by the multiple method with the same criterion (SCA5) for the first two components. The results obtained by the multiple method with maximal variance (SCA6) are more accurate than those obtained by the single method with maximal variance (SCA4). These observations are the same as those for the simple intermediate structure.

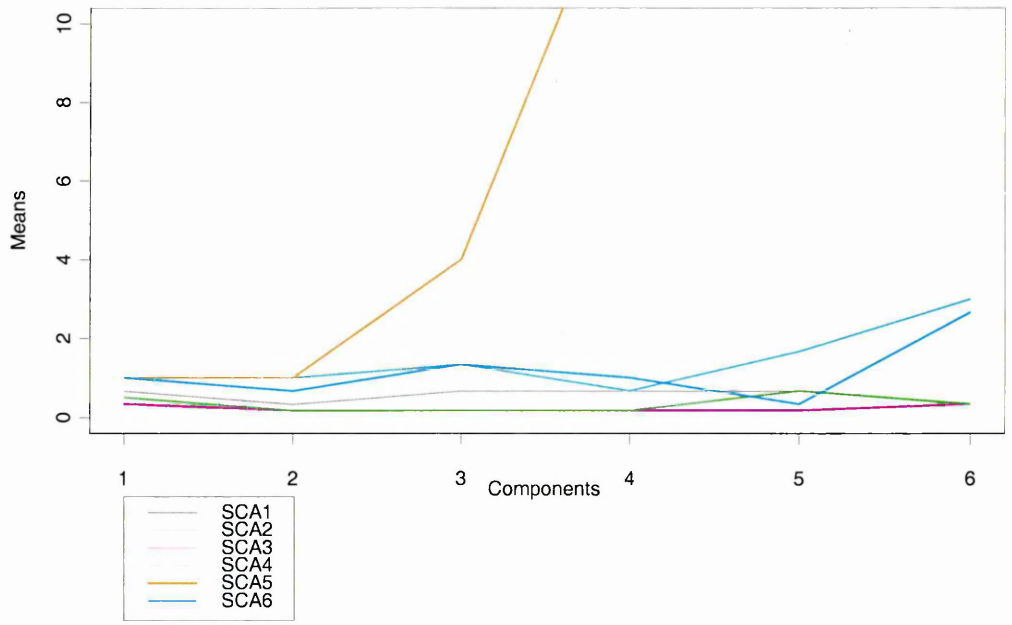

Figure 4.12: The simplicity of components produced by different SCA methods for 6 dimensional complex intermediate structure

SCA1 took five iterations (5 steps). SCA2 took eight iterations (8 steps). SCA3 took one iteration (1 step). SCA4 took two iterations (2 steps). SCA5 took eight iterations (15 steps). SCA2 took more steps than that taken by SCA1, SCA4 took more steps than that taken by SCA3. SCA6 took five iterations (7 steps). The number of steps taken by SCA5 is more than that taken by SCA2, the number of steps taken by SCA6 is more than that taken by SCA4.

The number of the steps taken by SCA5 is more than that taken by SCA6. Then considering the simplicity of the components (Figure 4.12), when $k = 0$, all the results obtained by the SCA methods are simple for the first two components. All the other observations about simplicity are the same as those for simple intermediate structure. This is to be expected in view of the steps of each SCA method taken because more steps generally mean less simple results.

SCA2 and SCA5 are the best methods for this structure given the tradeoff of accuracy and simplicity. This is the best result for this structure; actually the first two components obtained for the complex structure using SCA2 and SCA5 are the same as those for the simple intermediate structure using SCA2 and SCA5.

4.7 Discussion and conclusion

All in all, for our 6 dimensional data, SCA2, SCA5 and SCA6 seem to be the most accurate methods among all the SCA methods. For all the structures, the results obtained by SCA2 (single SCA method with maximal improvement in variance) are much more accurate than those obtained by SCA4 (single SCA method with maximal variance), the results obtained by SCA5 (multiple SCA method with maximal improvement in variance) are more accurate than those obtained by SCA6 (multiple SCA method with maximal variance). It seems that results obtained by SCA methods with maximal improvement in variance are usually more accurate than those obtained by SCA methods with maximal variance. In terms of accuracy, the maximal improvement in variance criterion is better than the maximal variance criterion. This is because, overall, the maximal improvement in variance criterion generally makes the variance of a simple component closer to the variance of the corresponding principal component compared to the maximal variance criterion. But the difference between the results obtained by SCA5 and SCA6 is not so obvious as the difference between the results obtained by SCA2 and SCA4. This is because the multiple SCA methods update as many pairs as possible in each iteration. So the multiple SCA methods generally update more than one pair at each iteration. Only the first pair have the maximal improvement in variance (or maximal variance), the second pair have the second maximal improvement in variance (or maximal variance), and so on. This process reduces

the difference between the maximal improvement in variance criterion and maximal variance criterion, so it reduces the difference between the results obtained by SCA5 and SCA6. For example the RI data in the first iteration (Section 3.3.5), SCA2 updates columns 2 and 4, SCA4 updates columns 1 and 3, SCA5 and SCA6 update all the four columns. So the transformation matrices of SCA2 and SCA4 are much different, but the transformation matrices of SCA5 and SCA6 are the same in the first iteration.

The restriction $l_r = l_s$ generally makes the calculation of SCA1 and SCA3 finish early, i.e, SCA1 generally took fewer steps than taken by SCA2 and SCA3 generally took fewer steps than taken by SCA4. So the results obtained by SCA1 and SCA3 generally are less accurate but more simple than those obtained by the corresponding method without the restriction $l_r = l_s$ (SCA1 with SCA2, SCA3 with SCA4). This is to be expected. At the start, all the columns of the transformation matrix have the same lengths, so the transformation matrices of SCA1 and SCA2 are the same and so are SCA3 and SCA4. But after a few steps, the lengths of all the columns are generally different because different columns are updated at different times or different value of b are obtained in the transformation matrix even if same pair of columns are updated. So when the improvement in variance (or variance) at each step is compared, the pairs with different lengths are not considered if the condition $l_r = l_s$ is imposed. But the maximal improvement in variance (or maximal variance) possibly is obtained within the directions with different lengths. This generally makes the results obtained by SCA2 more accurate than those obtained by SCA1 and the results obtained by SCA4 more accurate than those obtained by SCA3.

When $k = 0$, the results obtained by the single SCA methods with maximal improvement in variance (SCA2) generally are at least as accurate as those obtained by the multiple methods (SCA5) with maximal improvement in variance. This is possibly because for SCA5, at each iteration, the transformation matrix is the multiplication of the single updating matrices. Only the first single updating matrix corresponds to the maximal improvement in variance within all the pairs, the other single updating matrices correspond to the maximal improvement in variance within the pairs not chosen pairs in previous steps. This reduces the maximal improvement in variance. So the results obtained by SCA5 are generally not more accurate than those obtained

by SCA2.

The results obtained by the multiple method with maximal variance (SCA6) generally are more accurate than those obtained by the single method with maximal variance (SCA4). The reason is that the multiple method with maximal variance updates as many pairs as possible at each iteration, as said in the first paragraph of this section, this process reduces the difference between the maximal improvement in variance criterion and the maximal variance criterion, and the maximal improvement in variance criterion is better than the maximal variance criterion. So the results obtained by SCA6 are generally more accurate than those obtained by SCA4.

In general, multiple methods update more than one pair at each iteration. If updating one pair is thought to be similar to one iteration of single methods, and is called one step, multiple methods generally took more, or the same steps as those taken by corresponding single SCA methods. So generally the results obtained by single methods are more simple than those obtained the corresponding multiple methods (SCA2 with SCA5, SCA4 with SCA6). SCA methods with maximal variance generally took fewer steps than SCA methods with maximal improvement in variance. So the results obtained by SCA methods with maximal variance are generally more simple than those obtained by the corresponding SCA methods with maximal improvement in variance (SCA4 with SCA2, SCA6 with SCA5).

In general, SCA2, SCA5 and SCA6 are the best methods given the tradeoff of accuracy and simplicity of the components produced by the SCA methods. The best angles obtained by SCA methods for the complex structure are generally obtained if the SCA method got the corresponding simple structure back exactly, because generally the simple structure is the closest simple approximation of the corresponding complex structure.

SCA methods retrieved simple block and simple uniform structures. This is probably because the loadings of the transformation matrix at each step are more likely to be similar to the loadings of the simple block structure or simple uniform structure. For example, suppose the dimension of the data is 2, as said in Section 3.3.3, when $|b| \leq 1$, $B(b) = \begin{pmatrix} 2^k & 2^k b l_2^2 \\ 2^k b & -l_1^2 \end{pmatrix}$, when $k = 0$, the possible values of b are 0, -1 and 1. When $b = 0$, the transformation matrix is simple block structure. When $b = \pm 1$, the loadings of the transformation matrix is more likely to be smilar

to the loadings of the simple uniform structure at each step. For more than two dimensional data, the first two columns of multiplication (P in Section 3.3.3) of the transformation matrices are more likely to be simple block or simple uniform structures.

To sum up, generally the results obtained by SCA methods with maximal improvement in variance are at least as accurate as, but less simple, than those obtained by SCA methods with maximal variance, i.e. the results obtained by SCA1 are more accurate but less simple than those obtained by SCA3, the results obtained by SCA2 are more accurate but less simple than those obtained by SCA4, the results obtained by SCA5 are more accurate but less simple than those obtained by SCA6. The results obtained by the single SCA method with maximal improvement in variance are at least as accurate as and more simple than those obtained by the multiple SCA method with maximal improvement in variance, i.e. the results obtained by SCA2 are more accurate and more simple than those obtained by SCA5. However, the results obtained by multiple SCA method with maximal variance are more accurate but less simple than the results obtained by single SCA method with maximal variance, i.e. the results obtained by SCA6 are more accurate and less simple than those obtained by SCA4. The condition $l_r = l_s$ makes the results obtained by the SCA methods no more accurate but more simple than the results obtained by SCA methods without this condition, i.e. the results obtained by SCA1 are less accurate but more simple than those obtained by SCA2, the results obtained by SCA3 are less accurate but more simple than those obtained by SCA4.

Chapter 5

Results based on samples for 6 dimensional data when $k = 0$

5.1 Introduction

In Chapter 4, the population results applied to different variance covariance matrix V_0 showed that when $k = 0$, SCA methods are good for block and uniform structures. SCA methods are not very good for the intermediate structure. Generally SCA2, SCA5 and SCA6 are the best SCA methods.

In this chapter, I am going to check if all of the findings for the population results are true for sample simulation results. Here a simulation will be based on 500 data sets, each data set having 500 observations. Each data set will be assumed to have a normal distribution $N(0, V_0)$, where V_0 is different for different structure and V_0 is defined in Section 4.2.

In Section 5.2, I will first consider the sample simulation results for simple structures. As in Chapter 4, I just investigate the sample simulation results for the first two components of the simple structures because only the first two components have the features of each structure and they account for about 61% of the variance. The sample simulation results for other components could be investigated in the same way. The distribution of angles and the means of the first two components produced by PCA and SCA are given using boxplots separately, where the

definitions of angles and means are the same as those in Section 4.5.

In Section 5.3, I am going to consider the sample simulation results for complex structures, the distribution of angles and the means for the first two components produced by PCA and SCA will be displayed using boxplots separately.

It was noted that sometimes the results obtained by SCA methods can be improved greatly if the order of the components is not considered. For example, in Section 4.5.1, if the second component and fourth component obtained by SCA3 are exchanged, the results obtained by SCA3 are improved. i.e. if the angle between simple component i and actual component j is the smallest angle among all the angles between the actual component j and all the simple components in the structure, simple component i is taken as the approximation of actual component j. The sample simulation results ignoring order are obtained according to the following rules. The component with the smallest angle between the first actual component with all the simple components is chosen as the first component. If component i_1 is chosen as the first component ignoring the order of the components, the component with the smallest angle between the second actual component with all the components excluding component i_1 is chosen as the second component, and so on. It is expected that the sample simulation results after order of the components is ignored are improved. They should be at least as accurate as the results the order is not ignored because of the smallest angle between the actual component with all the simple components for the results ignored the order. So in Section 5.4, the sample simulation results obtained by SCA methods for 6 dimensional data if the order of the components is ignored are illustrated.

The last section, Section 5.5, given a discussion and conclusion.

5.2 Sample simulation results for simple structures

In this section and in the following two sections, the distribution of angles and the means of the first two components produced by SCA and PCA are displayed using boxplots. The order used is the first two components produced by PCA, followed by the first two components produced by SCA1 to the first two components produced by SCA6. In the following boxplots, different

colours are used. The black represents the difference between the median and the lower quartile, the blue represents the range between the median and the upper quartile. So that it is known exactly where the median of the angles of each SCA method is even if it equals one of the quartiles. So the difference between the results obtained by different SCA methods can be seen very clearly. In the following boxplots, the ends of the coloured sections are quartiles, anything further than 1.5 times of the inter-quartile range is considered an outlier, the horizontal lines are outliers, the horizontal brackets are the upper extreme and lower extreme excluding outliers.

5.2.1 Simple block structure

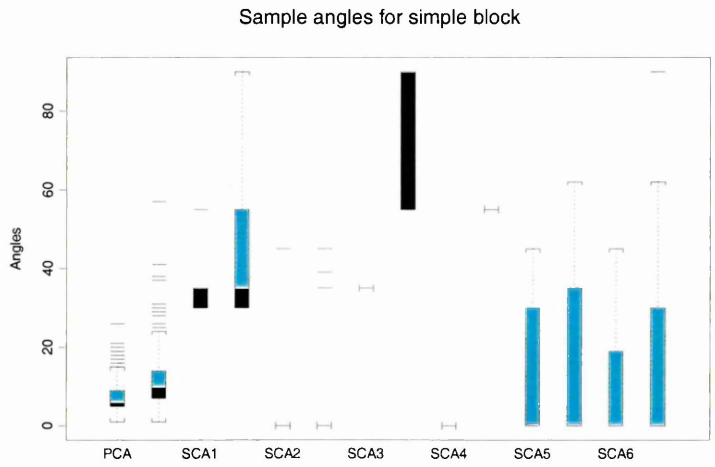

Figure 5.1: The angles between the first two original components and the corresponding components produced by PCA and SCA for 6 dimensional simple block structure

For the simple block structure in this simulation SCA2 is the most accurate method and SCA3 is the least accurate method among all the SCA methods (Figure 5.1). The results obtained by SCA2 are much more accurate than those obtained by PCA. This is clear in Figure 5.1, SCA2 only a few times did not get the right answers for components 1 and 2, i.e. SCA2 only a few times did not get component $(0, 0, 0, 1, 1, 1)$ for component 1, only a few times did not get $(-1, -1, -1, 0, 0, 0)$ for component 2. In other words, one line at anything other than zero for

SCA2 for component 1, and only three-non-zero lines for SCA2 for component 2. The medians of the angles obtained by SCA5 and SCA6 are zero for the first two components, SCA5 about 65% of the time got the first two components back, SCA6 almost 75% of the time got the first two components back. So SCA6 more often gets the exact results compared to SCA5. Furthermore the inter-quartile ranges obtained by SCA5 are longer than those obtained by SCA6. So the results obtained by SCA6 are more accurate than those obtained by SCA5. SCA4 gets the first component back exactly every time, but the angles for the second component are too big. The condition $l_r = l_s$ makes the results obtained by SCA methods less accurate than those obtained by corresponding SCA methods without this condition, i.e. the results obtained by SCA1 and SCA3 are less accurate than those obtained by other SCA methods.

The conclusions obtained here are similar to what happened when the SCA methods are applied directly to V_0 for the simple block structure. The results obtained by the SCA methods with maximal improvement in variance are more accurate than those obtained by SCA methods with maximal variance, i.e. the results obtained by SCA2 are more accurate than those obtained by SCA4. The results obtained by single SCA methods with maximal improvement in variance are more accurate than those obtained by the multiple method with maximal improvement in variance, i.e. the results obtained by SCA2 are more accurate than those obtained by SCA5. The last two conclusions are different from those obtained in Section 4.5.1. There SCA2, SCA5 and SCA6 retrieved the simple block structure. Notice that the medians for the first two components of SCA2, SCA5 and SCA6 are 0. This was expected considering the population results in Section 4.5.1. The sample simulation results of SCA2, SCA5 and SCA6 just indicate SCA2 is more stable than SCA5 and SCA6.

All the results obtained by SCA methods are very simple because all the means for the first two components are less than 10 (Figure 5.2). The actual absolute means of the original vectors are 0.5, 0.5, 0.3, 0.7, 0.7, 0.3 respectively. As the medians of the means for the first two components obtained by SCA1 are less than 0.5, they are too simple. The means obtained by SCA2 are almost all the same as the actual value for the first two components, so they are as simple as the actual components. The means obtained by SCA3 for the first two components are the smallest, so SCA3 is the simplest method, but the results obtained by SCA3 are usually

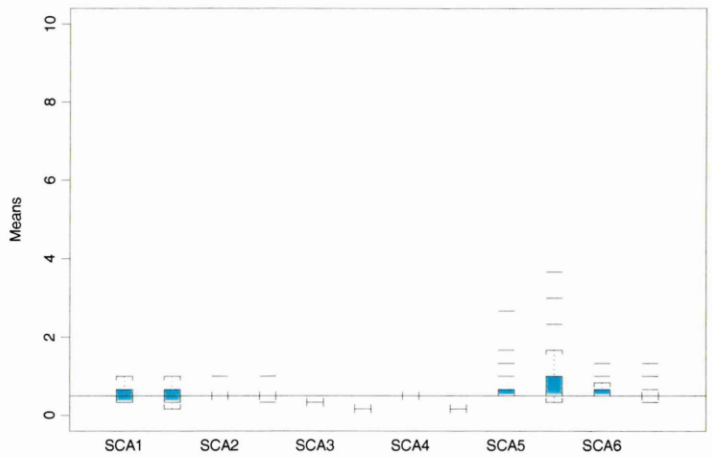

Figure 5.2: The means of the first two components produced by SCA for 6 dimensional simple block structure

too simple. The mean for the first component obtained by SCA4 is the same as the actual value, the mean for the second component obtained by SCA4 is smaller than the actual value, so the second component obtained by SCA4 is more simple than the actual component. The results obtained by SCA5 and SCA6 are generally slightly less simple than the actual value of the first two components.

The results obtained by SCA methods with maximal variance are more simple than those obtained by corresponding SCA methods with maximal improvement in variance, i.e. the results obtained by SCA3 are more simple than those obtained by SCA1, the results obtained by SCA4 are more simple than those obtained by SCA2, the results obtained by SCA6 are more simple than those obtained by SCA5. The results obtained by single SCA methods are more simple than those obtained by corresponding multiple SCA methods, i.e. the results obtained by SCA2 are more simple than those obtained by SCA5, the results obtained by SCA4 are more simple than those obtained by SCA6. This is similar to what was obtained when the methods are directly applied to V_0.

SCA2 is the most accurate method and the simplicity of the components obtained by SCA2 is the same as the actual value, so it is the best method for this structure.

5.2.2 Simple uniform structure

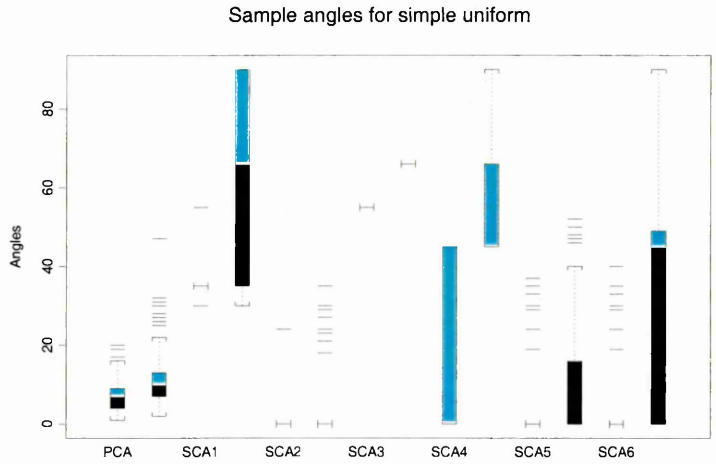

Figure 5.3: The angles between the first two original components and the first two components produced by PCA and SCA for 6 dimensional simple uniform structure

For the simple uniform structure, the results obtained by SCA2 (Figure 5.3) are more accurate than those obtained by PCA and other SCA methods. Only once did SCA2 not get the exact result for the first component, and only a few times did it not get the second component back exactly. This is expected considering the population results in Figure 4.3 because only SCA2 retrieved the simple uniform structure. The median of the angles obtained by SCA4 for the first component is zero, but the median of angles obtained by SCA4 for the second component is disappointing. The medians of the angles obtained by SCA4 are smaller than the population results for SCA4 (Figure 4.3). SCA5 and SCA6 only a few times did not get the first component. The median angle obtained by SCA5 for the second component is more than that of PCA, but the results obtained by SCA5 for the second component are good. The angle obtained by SCA6 for the second component is disappointing. Like the conclusion obtained in Section

4.5.2, the results obtained by SCA methods with maximal improvement in variance are more accurate than those obtained by corresponding SCA methods with maximal variance, i.e. the results obtained by SCA2 are more accurate than those obtained by SCA4, the results obtained by SCA5 are more accurate than those obtained by SCA6. The condition $l_r = l_s$ makes the results obtained by SCA methods less accurate than those obtained by the corresponding SCA methods which do not have this condition, i.e. the results obtained by SCA1 and SCA3 are less accurate than those obtained by other SCA methods. The results obtained by single updating with maximal improvement in variance (SCA2) are more accurate than those obtained by the corresponding multiple method (SCA5). The results obtained by single updating with maximal variance (SCA4) are less accurate than those obtained by the corresponding multiple method (SCA6).

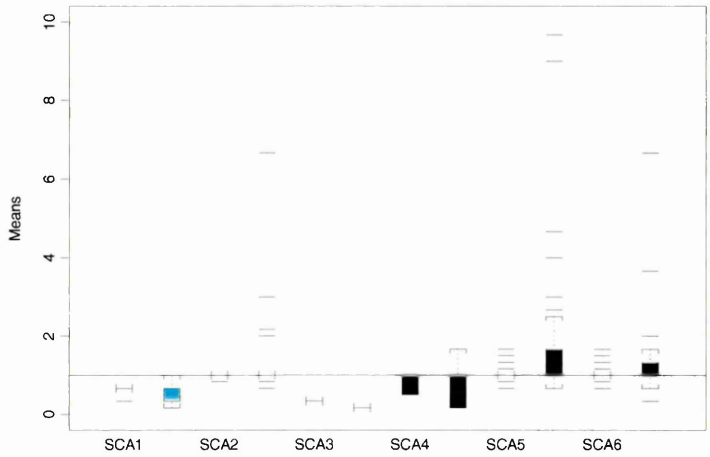

Figure 5.4: The means of the first two components produced by SCA for 6 dimensional simple uniform structure

Almost all the results obtained by SCA methods are simple because almost all the means of the components are less than 10 (Figure 5.4). For the first two components, the means of components obtained by SCA3 are the smallest, so it is the simplest method. Then is SCA1,

SCA4. The actual value of the means of components for this structure is 1, 1, 0.3, 0.7, 0.7, 0.3. The medians of the means for the first two components obtained by SCA2 are the same as the actual value. This is expected. SCA2 did not get the correct answer only a few times for the first two components. The results obtained by SCA1 and SCA3 are more simple than the actual components, so they are too simple. SCA1 on average took 4 steps, SCA2 on average took 7 steps. SCA3 generally took 1 steps and SCA4 generally took 2 steps. The fewer steps the method took, the more simple the results of the method generally. For the first component, the medians of the means of components obtained by SCA5 and SCA6 are the same as the actual value. For the second component, the medians of the means of components obtained by SCA5 and SCA6 are larger than the actual value, so the results obtained by SCA5 and SCA6 are not as simple as the actual component. The results obtained by SCA methods with maximal variance are more simple than those obtained by SCA methods with maximal improvement in variance. All the conclusions are the same as those obtained in Section 4.5.2 about the simplicity of the results. SCA2 is better than SCA5, SCA6 is better than SCA4.

Only rarely did SCA2 failed to retrieve the first two components, and the means of the first two components obtained by SCA2 are the same as the actual values of the first two components; it is the best method for simple uniform structure.

5.2.3 Simple intermediate structure

For the simple intermediate structure, the results obtained by SCA methods are much less accurate than the results in Section 5.2.1 and Section 5.2.2. Even for the best results, obtained by SCA2, the angles are large (Figure 5.5). This is not surprising because in Section 4.6.3 no method got the first two components back exactly even when applied to the variance covariance matrix V_0. SCA2 got the most accurate results for the first two components, SCA5 got the second accurate results. SCA3 was the least accurate method. The results obtained by SCA methods with maximal improvement in variance are more accurate than those obtained by corresponding SCA methods with maximal variance (SCA2 with SCA4, SCA5 with SCA6). The condition $l_r = l_s$ makes the results obtained by SCA1 less accurate than those obtained by SCA2 and the results obtained by SCA3 less accurate than those obtained by SCA4. The results obtained

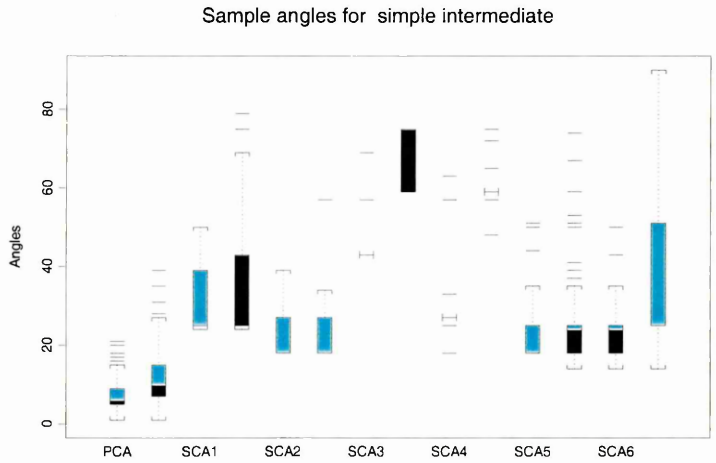

Figure 5.5: The angles between the first two original components and the first two components produced by PCA and SCA for 6 dimensional simple intermediate structure

by the single SCA method with maximal improvement in variance (SCA2) are more accurate than those obtained by the multiple method with maximal improvement in variance (SCA5). However the results obtained by the multiple method with maximal variance (SCA6) are more accurate than those obtained by single SCA method with maximal variance (SCA5). This is the same as what was found in Section 4.5.3.

SCA3 produces the simplest results (Figure 5.6). SCA4 is the second simplest method. The actual value of means of components for this structure is 1.5, 1.5, 1.3, 1, 1.5, 1 respectively. Almost all the means of the first two components produced by SCA methods are less than the actual value, so the first two components obtained by all the SCA methods are more simple than the actual components for the first two components. All the conclusions are the same as those in Section 4.5.3 about the simplicity of the results.

SCA methods did not perform very well for simple intermediate structure but SCA2 is the best. This is the same as the population results given in Section 4.5.3.

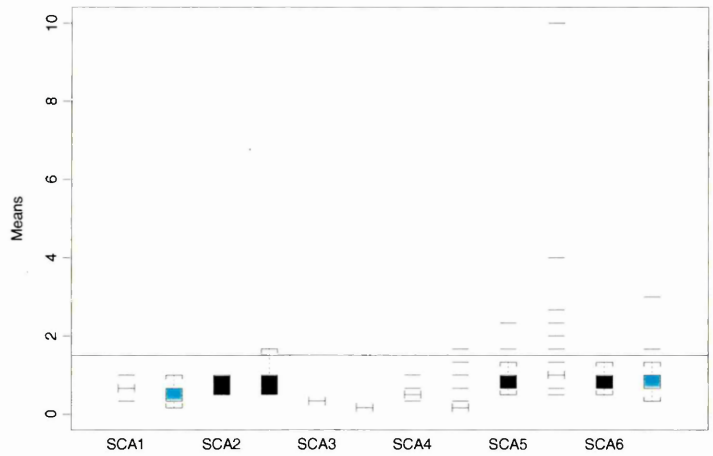

Figure 5.6: The means of the first two components produced by SCA for 6 dimensional simple intermediate structure

5.3 Sample simulation results - complex structures, $k = 0$

In this section, the sample simulation results for complex structures will be investigated.

5.3.1 Complex block structure

For complex block structure, SCA2 is the most accurate method (Figure 5.7). Comparing the results in Figure 5.7 with the results for the simple block structure in Figure 5.1, shows that the results obtained by SCA methods for complex block structure are less accurate than those obtained by SCA methods for the simple block structure. In particular, no SCA methods retrieved the complex structure. This is expected. The elements of loadings vectors obtained by SCA methods have to be smaller-value integers, i.e. the integers are less than 10. Checking the results of SCA2 for the simple block structure and complex block structure, SCA2 gets the most accurate results for simple and complex block structures. Also the results for simple structure

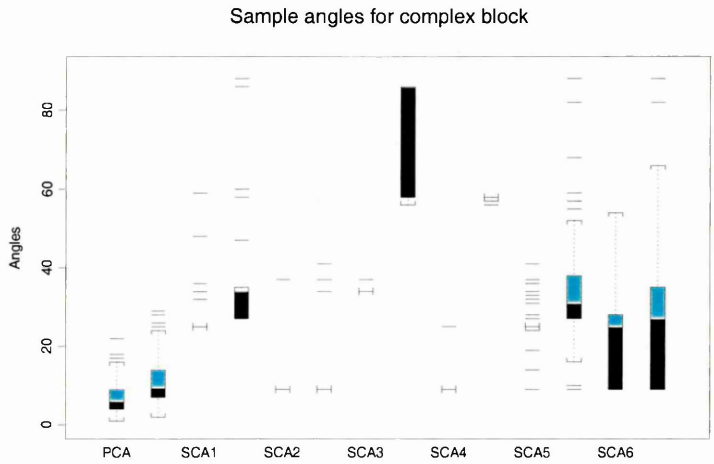

Figure 5.7: The angles between the first two original components and the first two components produced by PCA and SCA for 6 dimensional complex block structure

and complex structure are almost always the same as for simple block structure, and are the same as the population results for simple block and complex block structures (Figures 4.1 and 4.7). So the best results are the same as the angles $(9, 9, 16, 17, 13, 5)$ between simple block structure and complex block structure. SCA2 only once failed to retrieve the first component of the simple block structure, and only three times failed to retrieve the second component. The median angles obtained by SCA5 and SCA6 are similar for the first component. The results of single updating with maximal improvement in variance (SCA2) are generally more accurate than those obtained by the multiple updating with maximal improvement in variance (SCA5). The results obtained by the single methods with maximal improvement in variance are generally more accurate than those obtained by the corresponding single methods with maximal variance, i.e. the results obtained by SCA2 are more accurate than those obtained by SCA4, the results obtained by SCA1 are more accurate than those obtained by SCA3. These conclusions are the same as those for the simple block structure in Section 5.2.1.

All the results obtained by SCA methods are simple (Figure 5.8), i.e. all means of the first two components produced by SCA methods are less than 10. All the conclusions about the

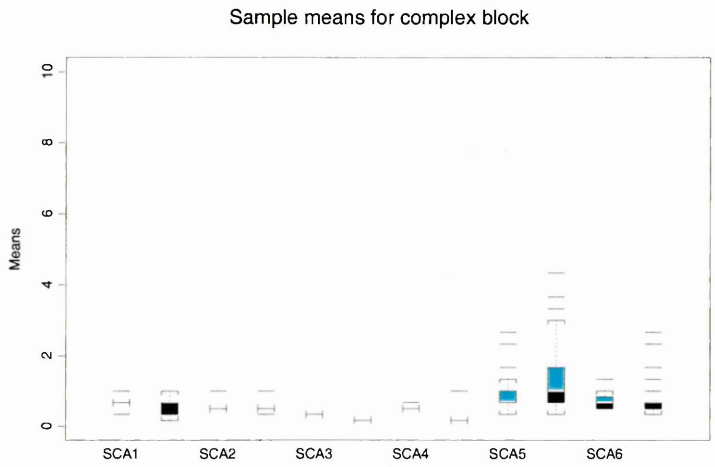

Figure 5.8: The means of the first two components produced by SCA for 6 dimensional complex block structure

simplicity are the same as those in simple block structure in Section 5.2.1.

Combining the accuracy and simplicity, SCA2 is the best method for complex block structure.

5.3.2 Complex uniform structure

For complex uniform structure, SCA2 and SCA5 generally got the same result for the first component (Figure 5.9), and they tended to get the best results for the first component. Only a few times did SCA2 and SCA5 not get the best results. SCA5 got the best results for the second component. So overall SCA5 is the most accurate method. The results obtained by SCA1 and SCA3 are very disappointing. This is expected from the population results for complex uniform structure in Section 4.6.2 (Figure 4.9). For this structure, the results obtained by SCA methods with maximal improvement in variance (SCA2 and SCA5) are generally more accurate than those obtained by other SCA methods. The results obtained by SCA5 and SCA6 are generally more accurate than those obtained by SCA5 and SCA6 in the complex block structure. The condition $l_r = l_s$ makes the results obtained by SCA1 and SCA3 less accurate than those obtained by other SCA methods. All the findings are the same as those for the simple uniform

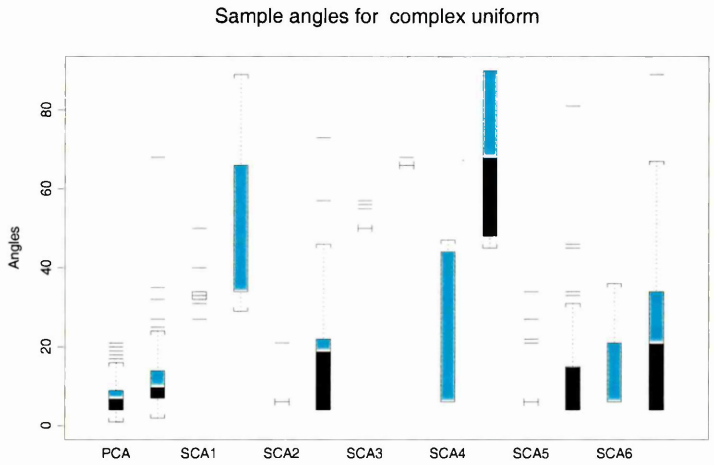

Figure 5.9: The angles between the first two original components and the first two components produced by PCA and SCA for 6 dimensional complex uniform structure

structure given in Section 5.2.2.

All the results obtained by SCA methods are simple. SCA3 is the simplest method (Figure 5.10), followed by SCA1. The results obtained by SCA1 are more simple than those obtained by SCA2, the results obtained by SCA3 are more simple than those obtained by SCA4. The results obtained by SCA5 and SCA6 are very similar. So all the conclusions about the simplicity of the results obtained by SCA methods are the same as those obtained in Section 4.6.2 for the complex uniform structure.

Combining accuracy with simplicity, SCA5 and SCA2 are the best methods, next is SCA6.

5.3.3 Complex intermediate structure

For complex intermediate structure, comparing the result in Figure 5.11 with the result in simple intermediate structure, all the methods are disappointing in terms of recovering the complex intermediate structure. However, the results obtained by all the SCA methods are less accurate than those obtained by PCA. No method got the first two components of corresponding simple intermediate structure back. But this is hardly surprising given what happened with V_0 in

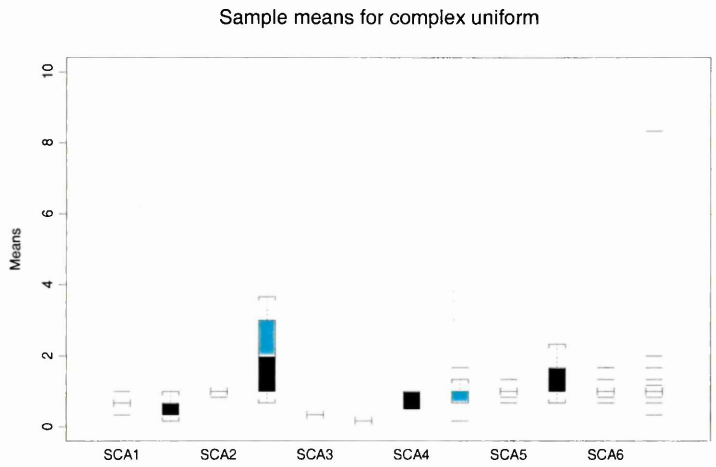

Figure 5.10: The means of the first two components produced by SCA for 6 dimensional complex uniform structure

Section 4.6.3. Even in that ideal case, i.e the case applied to the variance covariance V_0 itself, no SCA methods got the first two components of corresponding simple intermediate structure back exactly.

Comparing the SCA methods, SCA2 is the most accurate method, next is SCA5, SCA3 is the least accurate method. As for the population results for simple intermediate structure and complex intermediate structure (Figures 4.5 and 4.11), for the sample simulation results, the difference (Figures 5.5 and 5.11) between the angles of SCA2 for the complex intermediate structure and simple intermediate structure are almost the angles between the first two components of simple intermediate structure and complex intermediate structure (3° and 4°). The condition $l_r = l_s$ makes the results obtained by SCA methods less accurate than corresponding SCA methods without this condition (SCA1 with SCA2, SCA3 with SCA4). The results obtained by SCA2 are more accurate than those obtained by SCA4. The results obtained by SCA5 are more accurate than those obtained by SCA6, i.e. the results obtained by SCA methods with maximal improvement in variance are more accurate than those obtained by corresponding SCA methods with maximal variance. The results obtained by SCA2 are slightly more accurate than

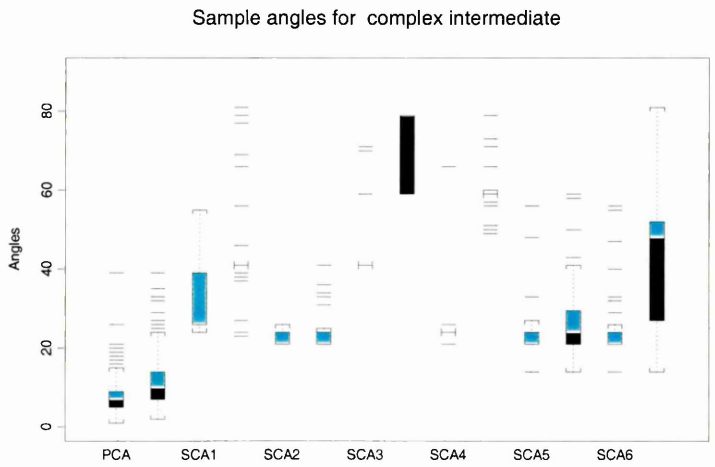

Figure 5.11: The angles between the first two original components and the first two components produced by PCA and SCA for 6 dimensional complex intermediate structure

those obtained by SCA5, the results obtained by SCA6 are more accurate than those obtained by SCA4. These are the same as what have been obtained for simple intermediate structure in Section 5.2.3.

All the results obtained by SCA methods (Figure 5.12) are simple, SCA3 is the simplest method. The condition $l_r = l_s$ makes the results obtained by SCA methods more simple than those obtained by the corresponding SCA methods without this condition. The results obtained by single methods are slightly more simple than those obtained by multiple methods, this is expected because SCA1 generally took 6 steps, SCA2 generally took 8 steps, SCA3 generally took 1 step, SCA4 generally took 5 steps, SCA5 generally took 15 steps and SCA6 generally took 8 steps, i.e. the more steps taken generally means less simple components produced. The results obtained by SCA methods with maximal variance are more simple than those obtained by SCA methods with maximal improvement in variance. In other words, the results obtained by SCA3 are more simple than those obtained by SCA2, the results obtained by SCA4 are more simple than those obtained by SCA2, the results obtained by SCA6 are more simple than those obtained by SCA5.

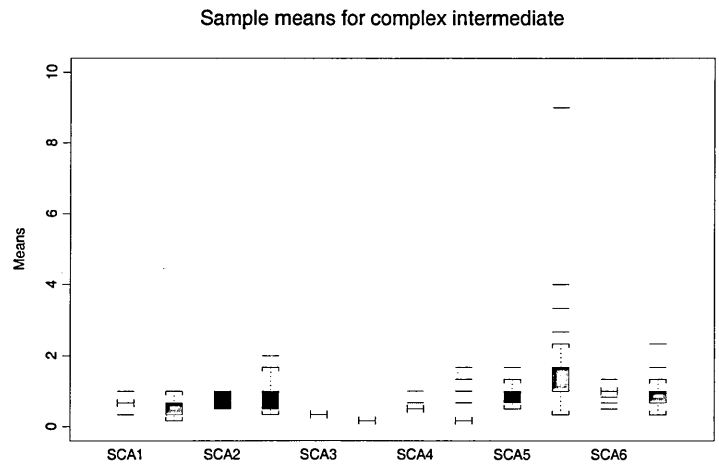

Figure 5.12: The means of the first two components produced by SCA for 6 dimensional complex intermediate structure

For this structure, when $k = 0$ no SCA methods got the first two components of the corresponding simple structure back.

Combining accuracy with simplicity, SCA2 is the best method for the complex intermediate structure.

5.4 Sample simulation results for 6 dimensional data if the order is ignored

In this section, I am going to investigate the sample simulation results for 6 dimensional data if the order is ignored. The results obtained by SCA1 and SCA3 will not be displayed because the angles ignoring the order of the components obtained by SCA1 and SCA3 are too large. Only the accuracy of the sample simulation results for 6 dimensional data when the order of components is ignored for SCA2, SCA4, SCA5 and SCA6 is investigated because almost all the results obtained by SCA methods are very simple for all the structures, the best SCA methods should be the methods with most accurate results.

Generally, the results ignoring the order are at least as accurate as the results before. This is what was expected because the results when ignoring the order of components are the minimal angles between the actual component and all the simple components (Section 5.1). I am going to consider the results ignoring the order of components for the simple and complex uniform structures. The results ignoring the order of the components for other structures are similar.

5.4.1 Simple uniform structure

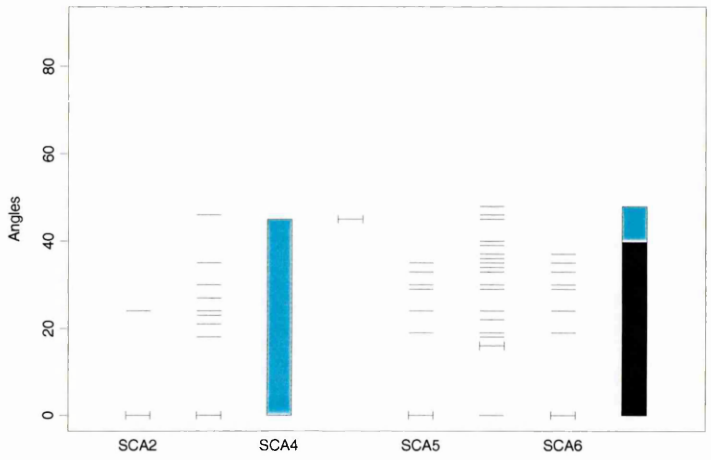

Figure 5.13: the results ignoring order for 6 dimensional simple uniform structure

For simple uniform structure, comparing this results (Figure 5.14) with those in Figure 5.3, the results obtained by SCA2 are the same. The results obtained by SCA5 and SCA6 are very similar for the first component, the results obtained by SCA5 and SCA6 being a little bit improved for the second component. This indicates that the components obtained by these three methods are generally in the right order. In other words, for simple uniform structure, component i has the minimal angle with actual component j, where i is equal to j. Actually, for this structure the results obtained by SCA2 are always in the right order and the results obtained by SCA5 and SCA6 are always in the right order for the first component. However

in about 10% of cases the second component produced by SCA5 are not closest to the second component of the simple uniform structure, and about 26% the second component produced by SCA6 are not closest to the second component of the simple uniform structure. This is very surprising. The explanation is in the small differences between the angles when ignoring the order of the components and the angles sticking to the order. The results obtained by SCA4 are similar for the first component, the results obtained by SCA4 are improved for the second component. The median of the angles obtained by SCA4 for the second component is the same, SCA4 gets the median of the angles for the second component all the time. However the results obtained by SCA4 are not so good as the results obtained by SCA2, so I don't give the details about how many times generally the results obtained by SCA4 are in the wrong order.

The improvements of the results obtained by SCA4 and SCA6 did not change the conclusions. All the conclusions are the same as those in Section 5.2.2. The results obtained by SCA methods with maximal improvement in variance are more accurate than those obtained by SCA methods with maximal variance, i.e. the results obtained by SCA2 are more accurate than those obtained by SCA4, the results obtained by SCA5 are more accurate than those obtained by SCA6. The results obtained by the multiple method with maximal variance (SCA6) are more accurate than those obtained by the single SCA method with maximal variance (SCA4).

5.4.2 Complex uniform structure

For complex uniform structure, comparing the results (Figure 5.14) with those in Figure 5.9, the results obtained by SCA2, SCA5 and SCA6 are similar. The results obtained by SCA2 and SCA5 and the results for the first component obtained by SCA6 are not in the right order only a few times (less than 3). The results obtained by SCA6 for the second component are not in the right order about 5% of times. The results ignoring the order obtained by SCA4 for the first component are similar to the previous results. For the second component obtained by SCA4 the results are improved. As I said in the previous subsection, because the results obtained by SCA4 are still not good compared to the results obtained by SCA2, SCA5 and SCA6. So I am not interested in how many times the results of SCA4 are not in the right order. The general findings about the accuratcy of the results obtained by SCA methods are the same as those for

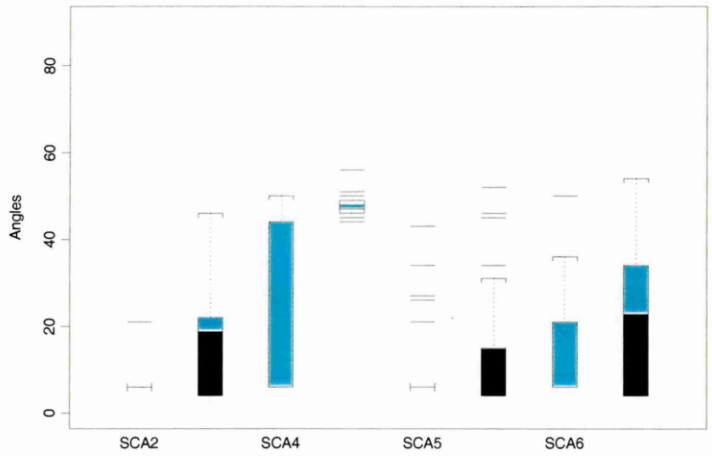

Figure 5.14: the results ignoring order for 6 dimensional complex uniform structure

complex uniform structure in Section 5.3.2.

5.5 Discussion and conclusion

Back to the results in Chapters 4 and 5, the conclusions are the same about which SCA method produces results that are the most accurate and which SCA method produces the results which are the most simple. In general, SCA2 gets the best results no matter which structure the SCA method is applied to.

The reason has been explained in Chapter 4, so only the conclusions are given here. For 6 dimensional data sets, when $k = 0$, it has been shown that the simple components technique is capable of recovering block structure and uniform structure. For example, SCA2 can recover simple block structure and simple uniform structure. The SCA technique is not very good for intermediate structure. SCA3 and SCA4 tend to retrieve simple block structure.

In general, the results obtained by the SCA methods with maximal improvement in variance are at least as accurate but not as simple as the results obtained by the SCA methods with maximal variance, i.e. the results obtained by SCA1 are generally more accurate but not as

simple as those obtained by SCA3, the results obtained by SCA2 are generally more accurate but not as simple as those obtained by SCA4, the results obtained by SCA5 are generally more accurate but not as simple as those obtained by SCA6. The results obtained by the single SCA method with maximal improvement in variance (SCA2) are at least as accurate as and more simple than those obtained by the multiple SCA method with maximal improvement in variance (SCA5). The results obtained by the single SCA method with maximal variance (SCA4) are less accurate but more simple than those obtained by the multiple method with maximal variance (SCA6). The condition $l_r = l_s$ generally makes results of SCA methods less accurate, but more simple than those obtained by other SCA methods. So the results obtained by SCA methods with this condition generally are easy to interpret but less accurate. In other words, the results obtained by SCA1 and SCA3 are no more accurate but more simple than those of the SCA methods without these restrictions. All of these conclusions are the same as those obtained in Section 4.7. Combining simplicity with accuracy of the components produced by SCA methods, generally SCA2, SCA5 and SCA6 are the best methods among all the SCA methods.

If the order of the components is not considered, the sample simulation results obtained by SCA methods for 6 dimensional data are improved. This is as expected because the results ignoring the order of the components are the minimal angle between the actual component and all the simple components. However ignoring the order of components does not change the general conclusions about the accuracy and simplicity of the components produced by SCA methods given in the previous paragraph.

Generally, for simple block and simple uniform structures, the most accurate sample simulation results obtained by SCA method were more accurate than the sample simulation results obtained by PCA. This is surprising at first glance because I suppose the easier interpretation of simple components should sacrifice some of the accuracy of the results. However if the normalizing constant for simple structures is omitted, the loadings of the simple block and simple uniform structures are integers. But for PCA, the loadings of the principal components generally are not integers for sample simulation results. Furthermore as said in Section 4.7, the loadings of the transformation matrix at each step are more likely to be similar to the loadings of the simple block structure or simple uniform structure when $k = 0$. So SCA methods get more

accurate sample simulation results than those obtained by PCA for simple block and simple uniform structures.

In contrast to the simple block and simple uniform structures, for complex block and complex uniform structures, the most accurate sample simulation results obtained by a SCA method were slightly less accurate than the sample simulation results obtained by PCA but the components produced by the SCA methods were generally more simple to interpret than the components produced by PCA.

For simple intermediate and complex intermediate structures, the most accurate sample simulation results obtained by the SCA method were less accurate than the sample simulation results obtained by PCA but the interpretation of the components produced by the SCA methods is generally easier than that of the components produced by PCA.

The next chapter investigates whether the findings in Chapters 4 and 5 for the block, uniform and intermediate structures in 6 dimensions are also true for them in 8 and 10 dimensions.

Chapter 6

Population results for large dimensional data: 8 and 10 dimensional data

6.1 Introduction

In Chapters 4 and 5, the performance of the SCA methods when applied to 6 dimensional data were investigated, and some conclusions about the accuracy and simplicity of the components produced by SCA methods were obtained. It is important to know if all the conclusions in Chapters 4 and 5 also apply to large dimensional data. So in this chapter, the population results for 8 and 10 dimensional data will be investigated. Although the results obtained by SCA1 and SCA3 are disappointing in Chapters 4 and 5, all the six SCA methods are checked in this chapter again. This is because I am not sure whether SCA1 and SCA3 are just disappointing for 6 dimensional data or generally disappointing whatever the dimension of the data. Again the three structures are investigated, block structure, uniform structure and intermediate structure. Each structure will have both a simple form and complex form. As in Chapter 4, the conclusions here are just based on the first two components of the structures, because generally only the first two components of the 8 and 10 dimensional data have the feature of the structures and

very similar to the first two components of 6 dimensional data (the main difference being the dimensionality). As in Section 4.4, angle is defined as the angle between the simple component and the corresponding original eigenvector, mean is defined as the mean of the absolute values of the loadings in a simple component, and the angles and means are used to measure the accuracy and simplicity of the results obtained by SCA methods respectively.

In Section 6.2, the data used in this chapter is given. In Section 6.3, the population results for 8 dimensional data are considered. Section 6.4 gives the population results for 10 dimensional data. The last section, Section 6.5, is the discussion and conclusion.

6.2 Data used in this chapter

As for the 6 dimensional data, the eigenvalues used for large dimensional data are all different, and none of them are zero. In the following examples, for all 8 dimensional data, λ is (20, 16, 14, 12, 10, 6, 4, 2), the first two components explained 43% of the variance. For all 10 dimensional data, λ is (20, 18, 16, 14, 12, 10, 8, 6, 4, 2), the first two components explained 35% of the variance. So the corresponding eigenvectors are uniquely defined. The 8 and 10 dimensional eigenvectors matrices are listed below.

6.2.1 8 Dimensional data

As in Section 4.3, the normalizing constants for the 8 dimensional simple structures are omitted.

The 8 dimensional simple block structure used was

$$A = \begin{pmatrix} 0 & -1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & -1 & 0 & 0 & -1 & 1 & 0 & 1 \\ 0 & -1 & 0 & 0 & 0 & 1 & 0 & -2 \\ 0 & -1 & 0 & 0 & 0 & -3 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & -1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -3 & 0 \end{pmatrix}.$$

If columns 1, 3, 4 and 7 are put together, and other columns are put together, it is clear that this is a simple block structure. The first two components of 8 dimensional data are very similar to those of the 6 dimensional simple block structure.

The 8 dimensional simple uniform structure was taken to be

$$A = \begin{pmatrix} -1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ -1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 & -1 & 1 & -1 & 1 \end{pmatrix}.$$

The absolute values of all the elements for every components are the same, though this is not necessary, and all of the elements are integers, so this is a simple uniform structure. For the 6 dimensional simple uniform structure, only the first two components have the features of the uniform structure. In the 8 dimensional structure, all the components have the features of uniform structure, so this is a really uniform structure. For the first two components, 6 dimensional simple uniform structure and 8 dimensional simple uniform structure are very similar.

The 8 dimensional simple intermediate structure was taken to be

$$A = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & -2 & -1 & 0 & 0 & 0 & 1 & 0 \\ 1 & -2 & 0 & 1 & 0 & 0 & -1 & 0 \\ 1 & -2 & 0 & -1 & 0 & 0 & -1 & 0 \\ 2 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 2 & 1 & 0 & 0 & -1 & 0 & 0 & 1 \\ 2 & 1 & 0 & 0 & 0 & 1 & 0 & -1 \\ 2 & 1 & 0 & 0 & 0 & -1 & 0 & -1 \end{pmatrix}.$$

This structure is similar to the 6 dimensional simple intermediate structure, only the first two components have the features of the simple intermediate structures.

The following complex structures are obtained from the corresponding simple structure. The complex structure and the corresponding simple structure have the similar directions.

The 8 dimensional complex block structure was taken to be

$$A = \begin{pmatrix} 0.022 & -0.464 & -0.060 & 0.011 & 0.719 & 0.285 & -0.006 & 0.427 \\ 0.020 & -0.498 & 0.095 & 0.041 & -0.686 & 0.283 & 0.005 & 0.436 \\ 0.014 & -0.510 & -0.009 & -0.007 & 0.006 & 0.338 & -0.025 & -0.790 \\ 0.014 & -0.521 & 0.085 & 0.007 & 0.024 & -0.848 & 0.018 & -0.028 \\ 0.539 & -0.012 & -0.682 & 0.426 & -0.064 & -0.045 & 0.236 & -0.006 \\ 0.498 & 0.070 & 0.716 & 0.390 & 0.087 & 0.050 & 0.265 & -0.035 \\ 0.482 & -0.006 & -0.012 & -0.812 & -0.016 & 0.010 & 0.328 & 0.014 \\ 0.477 & 0.020 & 0.031 & -0.070 & -0.007 & -0.021 & -0.875 & 0.015 \end{pmatrix}.$$

The angles between the simple block structure and complex block structure are 4, 5, 9, 5, 7, 5, 4 and 4 degrees respectively.

The 8 dimensional complex uniform structure was taken to be

$$A = \begin{pmatrix} -0.334 & 0.460 & -0.253 & -0.397 & -0.347 & -0.286 & 0.420 & 0.274 \\ -0.376 & 0.245 & -0.283 & -0.431 & 0.300 & 0.383 & -0.388 & -0.381 \\ -0.334 & 0.362 & 0.277 & 0.408 & 0.356 & 0.436 & 0.237 & 0.374 \\ -0.398 & 0.301 & 0.282 & 0.420 & -0.316 & -0.374 & -0.373 & -0.341 \\ 0.352 & 0.335 & -0.418 & 0.374 & -0.297 & 0.349 & 0.272 & 0.405 \\ 0.315 & 0.369 & -0.423 & 0.194 & 0.334 & -0.340 & -0.424 & 0.374 \\ 0.329 & 0.379 & 0.406 & -0.219 & 0.441 & -0.299 & 0.304 & -0.398 \\ 0.379 & 0.337 & 0.425 & -0.296 & -0.411 & 0.338 & -0.301 & -0.246 \end{pmatrix}.$$

The angles between the simple uniform structure and complex uniform structure are 4, 10, 12, 14, 8, 7, 11 and 9 degrees respectively.

The 8 dimensional complex intermediate structure was taken to be

$$A = \begin{pmatrix} 0.203 & 0.480 & -0.682 & -0.004 & 0.011 & -0.006 & -0.512 & -0.005 \\ 0.188 & 0.416 & 0.730 & -0.043 & 0.012 & -0.018 & -0.507 & -0.016 \\ 0.191 & 0.518 & 0.042 & 0.663 & -0.006 & -0.046 & 0.501 & -0.015 \\ 0.179 & 0.428 & -0.001 & -0.745 & -0.010 & 0.038 & 0.478 & 0.003 \\ 0.442 & -0.176 & 0.000 & 0.008 & 0.709 & -0.055 & 0.021 & 0.517 \\ 0.495 & -0.193 & 0.001 & 0.013 & -0.703 & -0.045 & -0.005 & 0.470 \\ 0.387 & -0.183 & -0.016 & -0.042 & 0.028 & -0.743 & 0.017 & -0.511 \\ 0.515 & -0.207 & -0.001 & 0.042 & 0.043 & 0.662 & 0.009 & -0.500 \end{pmatrix}.$$

The angles between the simple intermediate structure and complex intermediate structure are 7, 6, 3, 5, 3, 6, 2 and 2 degrees respectively.

6.2.2 10 Dimensional data

As in Sections 4.3 and 6.2.1, the normalizing constants for the 10 dimensional simple structures are omitted. The 10 dimensional simple block was taken to be

$$A = \begin{pmatrix} 0 & -1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & 1 & 1 \\ 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & -2 & 1 \\ 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -3 \\ 0 & -1 & 0 & 0 & 0 & -4 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & -2 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -3 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & -4 & 0 & 0 \end{pmatrix}.$$

This is clearly a simple block structure because columns 1, 3, 4, 7 and 8 consist of one element, the other columns consist of other element. For the 10 dimensional simple block structure,

the first two components are very similar to those for the 8 and 6 dimensional simple block structures.

The 10 dimensional simple uniform structure was taken to be

$$A = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -4 \\ -1 & 1 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \\ -1 & 1 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \\ -1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ -1 & 1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & -1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & -1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & -4 & 0 & 0 & 0 \end{pmatrix}.$$

The absolute values of the loadings in the first two components are the same. So only the first two components have the features of uniform structure. These two components are very similar to those of 8 and 6 dimensional simple uniform structures.

The 10 dimensional simple intermediate structure was taken to be

$$A = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & -2 & -1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & -2 & 0 & 1 & 0 & 0 & -1 & 0 & 1 & 0 \\ 1 & -2 & 0 & -1 & 0 & 0 & -1 & 0 & 1 & 0 \\ 1 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & -4 & 0 \\ 2 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 2 & 1 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 1 \\ 2 & 1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 1 \\ 2 & 1 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 1 \\ 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -4 \end{pmatrix}.$$

The first two components have the feature of intermediate structure. These two components are very similar to those for 8 and 6 dimensional simple intermediate structures.

The 10 dimensional complex structures used were obtained from the corresponding simple structures. So, the directions of the complex form and the simple form are very close.

The specific complex block structure for the 10 dimensional data taken to be

$$A = \begin{pmatrix} 0.110 & -0.469 & 0.079 & 0.016 & -0.688 & -0.137 & -0.029 & -0.006 & 0.405 & 0.323 \\ 0.081 & -0.393 & 0.097 & 0.043 & 0.715 & -0.277 & 0.017 & -0.037 & 0.392 & 0.290 \\ 0.096 & -0.415 & 0.083 & 0.041 & -0.002 & -0.246 & 0.002 & 0.031 & -0.823 & 0.265 \\ 0.083 & -0.440 & 0.083 & 0.039 & -0.019 & -0.223 & 0.014 & 0.009 & 0.031 & -0.860 \\ 0.067 & -0.425 & 0.074 & 0.036 & 0.108 & 0.888 & 0.076 & -0.032 & -0.036 & -0.001 \\ 0.468 & -0.078 & -0.635 & -0.464 & 0.049 & 0.028 & -0.302 & 0.248 & 0.006 & -0.007 \\ 0.386 & 0.181 & 0.748 & -0.361 & 0.013 & 0.040 & -0.276 & 0.225 & 0.003 & -0.012 \\ 0.443 & 0.132 & -0.050 & 0.805 & 0.005 & 0.034 & -0.293 & 0.224 & 0.016 & -0.004 \\ 0.458 & 0.118 & -0.007 & -0.010 & -0.024 & -0.040 & 0.859 & 0.189 & 0.017 & 0.010 \\ 0.432 & 0.098 & -0.005 & -0.022 & -0.016 & -0.012 & -0.048 & -0.894 & -0.037 & -0.017 \end{pmatrix}.$$

The angles between the simple block structure and complex block structure are 12, 17, 12, 6, 7, 7, 6, 4, 4 and 3 degrees respectively.

The 10 dimensional complex uniform structure was taken to be

$$A = \begin{pmatrix} -0.280 & 0.323 & 0.002 & -0.018 & -0.009 & 0.007 & 0.005 & -0.006 & -0.045 & 0.902 \\ -0.281 & 0.336 & 0.696 & 0.038 & 0.040 & 0.519 & -0.056 & -0.013 & -0.045 & -0.214 \\ -0.278 & 0.315 & -0.694 & -0.198 & -0.044 & 0.505 & 0.010 & 0.016 & -0.049 & -0.209 \\ -0.285 & 0.353 & 0.108 & -0.635 & 0.008 & -0.572 & 0.043 & -0.018 & -0.018 & -0.224 \\ -0.297 & 0.384 & -0.130 & 0.744 & -0.012 & -0.381 & -0.011 & -0.034 & -0.023 & -0.214 \\ 0.396 & 0.292 & 0.005 & 0.003 & 0.021 & -0.030 & -0.129 & 0.655 & -0.557 & -0.004 \\ 0.396 & 0.258 & -0.022 & -0.024 & -0.004 & 0.006 & -0.188 & -0.750 & -0.422 & 0.005 \\ 0.298 & 0.307 & -0.057 & -0.016 & 0.658 & 0.008 & -0.319 & 0.046 & 0.525 & 0.018 \\ 0.319 & 0.307 & 0.035 & -0.014 & -0.744 & 0.012 & -0.170 & 0.057 & 0.466 & 0.006 \\ 0.302 & 0.265 & 0.027 & 0.031 & 0.098 & 0.051 & 0.901 & -0.036 & 0.104 & 0.000 \end{pmatrix}.$$

The angles between the simple uniform structure and complex uniform structure are 8, 7, 11, 13, 8, 9, 10, 7, 10 and 2 degrees respectively.

The 10 dimensional complex intermediate structure was taken to be

$$A = \begin{pmatrix} 0.209 & -0.375 & -0.682 & 0.127 & 0.021 & -0.027 & -0.497 & 0.010 & -0.295 & 0.010 \\ 0.174 & -0.356 & 0.722 & -0.020 & -0.034 & -0.072 & -0.464 & 0.008 & -0.316 & -0.007 \\ 0.237 & -0.382 & 0.073 & 0.656 & 0.162 & 0.064 & 0.553 & -0.034 & -0.160 & 0.016 \\ 0.207 & -0.455 & -0.078 & -0.722 & -0.076 & 0.028 & 0.439 & -0.003 & -0.151 & 0.005 \\ 0.187 & -0.417 & 0.031 & 0.033 & -0.001 & 0.009 & -0.160 & -0.012 & 0.874 & 0.007 \\ 0.383 & 0.185 & -0.009 & 0.082 & -0.670 & 0.003 & 0.021 & -0.550 & -0.002 & 0.244 \\ 0.377 & 0.229 & 0.025 & -0.154 & 0.713 & 0.088 & -0.074 & -0.449 & 0.012 & 0.241 \\ 0.382 & 0.192 & 0.029 & -0.009 & -0.090 & 0.673 & -0.049 & 0.546 & -0.002 & 0.235 \\ 0.428 & 0.209 & -0.013 & 0.001 & 0.013 & -0.726 & 0.073 & 0.443 & 0.034 & 0.206 \\ 0.417 & 0.198 & -0.003 & -0.012 & -0.009 & 0.036 & -0.001 & -0.026 & 0.013 & -0.886 \end{pmatrix}.$$

The angles between the simple intermediate structure and complex intermediate structure are 4, 5, 7, 13, 12, 8, 13, 6, 9 and 2 degrees respectively.

6.3 Population results for 8 dimensional data, $k = 0$

In this section, I will investigate the population results for 8 dimensional data. As for 6 dimensional data, first consider the population results for the 8 dimensional simple structures.

6.3.1 Simple structures - 8 dimensional data

For all the 8 dimensional simple structures, generally SCA1, SCA2, SCA5 and SCA6 got the most accurate results for the first two components. All of these methods except SCA6 use the maximal improvement in variance criterion. it is clearly that the results obtained by the SCA methods with maximal improvement in variance generally are at least as accurate as those obtained by the SCA methods with maximal variance. In terms of accuracy, the maximal improvement in variance criterion generally is better than the maximal variance criterion. This is the same as observed in Section 4.5 for 6 dimensional data.

The results obtained by all the SCA methods are simple, i.e. the means of the first two components produced by SCA methods are less than 10. For the first two components, SCA1,

SCA2, SCA5 and SCA6 got the same means because they got the same components. The components obtained by SCA3 and SCA4 generally are more simple than the actual components, so they are too simple generally. So the results of SCA1 are more simple than those by SCA2. The results obtained by SCA2 generally are more simple than those by SCA5. The results obtained by SCA4 are generally more simple than those taken by SCA6. The results obtained by SCA6 generally are more simple than those obtained by SCA5. As for 6 dimensional data in Section 4.7, again the general conclusions of the accuracy and simplicity of the components produced by SCA methods are the same for 8 dimensional simple structures. Generally the results obtained by SCA methods with maximal improvement in variance are at least as accurate but not as simple as those obtained by SCA methods with maximal variance. In other words, the results obtained by SCA1 are more accurate but not as simple as those obtained by SCA3, the results obtained by SCA2 are more accurate but not as simple as those obtained by SCA4, the results obtained by SCA5 are more accurate but not as simple as those obtained by SCA6. The results obtained by single SCA method with maximal improvement in variance generally are at least as accurate as and more simple than those obtained by multiple SCA method with maximal improvement in variance. For all the simple structures, the results obtained by SCA2 are the same and more simple than those obtained by SCA5. The results obtained by the multiple SCA method with maximal variance generally are more accurate but not as simple as the results obtained by the single SCA methods with maximal variance. In other words the results obtained by SCA6 are more accurate but not as simple as those obtained by SCA4. The condition $l_r = l_s$ makes the results obtained by the SCA methods no more accurate but more simple than the results obtained by SCA methods without this condition (SCA1 with SCA2, SCA3 with SCA4).

Due to every component of 8 dimensional simple uniform structure having the feature of uniform structure, only the simple uniform structure is investigated in more detail. SCA1, SCA2, SCA5 and SCA6 retrieved all the components exactly for the simple uniform structure (Figure 6.1). In order to distinguish these four methods, I add 2° and 4° for the results obtained by SCA6 and SCA1 respectively, minus 2° for the results obtained by SCA5. Recall that for the 6 dimensional simple uniform structure, only did SCA2 retrieved the simple uniform structure

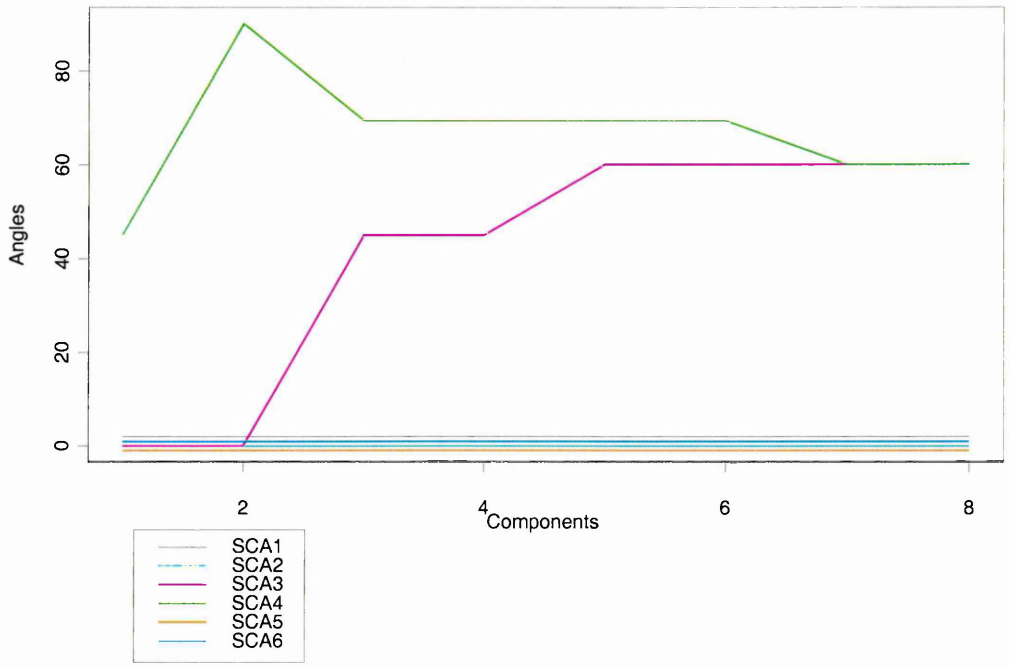

Figure 6.1: The accuracy of components produced by SCA methods for 8 dimensional simple uniform structure

exactly, so SCA1, SCA2, SCA5 and SCA6 are very good for 8 dimensional simple uniform structure. For 8 dimensional simple uniform structure, SCA3 retrieved the first two components exactly. SCA4 is the least accurate method.

Strangely the results obtained by SCA3 are more accurate than those obtained by SCA4 because generally the results obtained by SCA4 are more accurate than those obtained by SCA3. SCA3 and SCA4 use the same criterion, maximal variance, in each step, but SCA3 imposed the extra condition $l_r = l_s$. This makes SCA3 more likely to get the correct answer for this structure because all the columns of the simple uniform structure should have the same lengths. However, for this structure SCA3 took 6 steps, SCA4 only took 2 steps. Generally SCA4 took more steps than SCA3 took (Sections 4.5, 4.6, 5.2 and 5.3). In the first step, columns 1 and 2 (i.e. directions d_1 and d_2) are updated for both SCA3 and SCA4. In the second step, columns 1 and

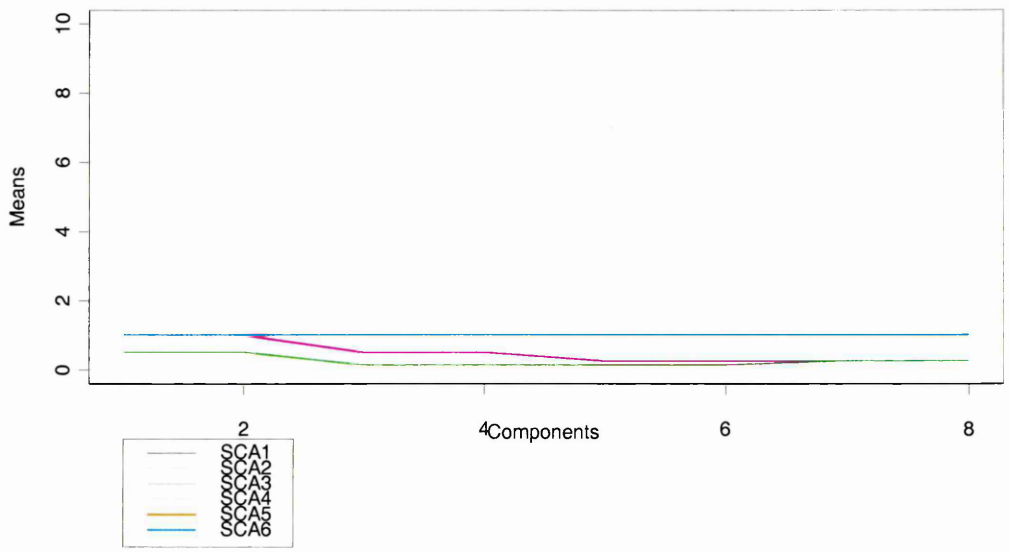

Figure 6.2: The simplicity of components produced by SCA methods for 8 dimensional simple uniform structure

3 are updated by SCA4, after this updating, the corresponding variance-covariance matrix V_2 for the 8 dimensional simple uniform structure is a diagonal matrix. So the calculation of SCA4 stopped. But SCA3 cannot update this pair because of the different lengths of columns 1 and 3 (column 1 has been updated twice but column 3 only updated once and b is not zero in the first 2 steps). In the second step, columns 3 and 4 are updated for SCA3 because by updating this pair the maximal variance is obtained among all the pairs with the same lengths. In the third step, the corresponding variance-covariance matrix V_3 is not a diagonal matrix for SCA3. So SCA3 took more steps to let the corresponding variance-covariance matrix become a diagonal matrix. For this structure, SCA1 and SCA2 took 12 iterations (12 steps), all the transformation matrices of SCA1 and SCA2 are the same in each iteration, and so end up with the same results.

For the 8 dimensional simple uniform structure, the updating pairs by the single SCA method

with the maximal improvement in variance just have the same length in each iteration. In the first four iterations, columns 1 and 2, 3 and 4, 5 and 6, 7 and 8 are updated respectively, all the columns are updated once and b is the same at each step, so all the columns have the same length. From iteration 5 to iteration 8, columns 1 and 3, 5 and 7, 2 and 4, 6 and 8 are updated respectively. So in iterations 5 to 8, all the columns are updated again and b is the same at each step, so they have the same lengths. From iteration 9 to iteration 12, columns 1 and 5, 2 and 6, 3 and 7, 4 and 8 are updated respectively, all the columns are updated and b is the same at each step, so they have the same lengths. That's why SCA1 and SCA2 get the same results.

From the columns updated each iterations of SCA1 and SCA2, it is not surprising that SCA5 gets the same results as those of SCA1 and SCA2, and that SCA5 only takes 3 iterations. All three iterations performed by SCA5 updates all the columns once. I am going to explain the first iteration of SCA5, the other two iterations can be interpreted in a similar way. Recall that SCA2 and SCA5 use the same criterion, maximal improvement in variance, the only difference is that SCA2 updates one pair and SCA5 updates as many pairs as possible at each iteration. Suppose that the transformation matrices of SCA2 in the first four iterations are P_1, P_2, P_3 and P_4 respectively. Noted that P_1, P_2, P_3 and P_4 update different columns, after updating columns 1 and 2, SCA5 get the maximal improvement in variance within all columns, and updating columns 3 and 4 SCA5 get the maximal improvement in variance within all the columns except columns 1 and 2, an so on. So $P_1P_2P_3P_4$ is the transformation matrix of SCA5 in iteration 1 because for 8 dimensional data SCA5 at most updates four pairs.

SCA5 and SCA6 have the same transformation matrix for all the three iterations, so SCA5 and SCA6 get the same results. SCA3 took 7 iterations (7 steps), SCA4 took 3 iterations (3 steps). SCA5 took 3 iterations (12 steps). SCA6 took 3 iterations (12 steps).

The results obtained by all the SCA methods (Figure 6.2) are very simple. The means of the first two components obtained by SCA1, SCA2, SCA3, SCA5 and SCA6 are the same as the actual value for the first two components. This is as expected because SCA1, SCA2, SCA3, SCA5 and SCA6 got the first two components of the simple uniform structure back exactly.

SCA1, SCA2, SCA5 and SCA6 got all the components back exactly, so these are the best methods for this structure.

6.3.2 Complex structures-8 dimensional data

For all the 8 dimensional complex structures, in general, SCA1, SCA2, SCA5 and SCA6 got the most accurate results for 8 dimensional complex structures. In order to distinguish these four methods, I add 2° and 4° for the results obtained by SCA6 and SCA1, minus 2° for the results obtained by SCA5. The results obtained by SCA1, SCA2, SCA5 and SCA6 for the complex structures are the same as those for the corresponding 8 dimensional simple structures. For complex structures, the differences between the best results for the complex structures and the simple structures are just the angles between the first two components of the complex structures and the first two components of the corresponding simple structures. SCA1, SCA2, SCA5 and SCA6 gave simple block and simple uniform structures. I am not surprised about this because the simple structure is generally the closest simple approximation of the corresponding complex structure by construction. SCA3 and SCA4 are the least accurate methods, in order to distinguish SCA3 and SCA4, I add 2o for the results of SCA4. The other conclusions about the accuracy and simplicity of the results for the SCA methods are the same as corresponding 8 dimensional simple structures (the first paragraph of Section 6.3.1). This is expected because all the SCA methods get the same components for the first two components for both simple structure and corresponding complex structure. Take 8 dimensional complex uniform structure as an example.

For 8 dimensional complex uniform structure (Figure 6.3), SCA1, SCA2, SCA5 and SCA6 obtained the most accurate results for the first two components, the results are just the angles between the first two components of the complex uniform structure and the first two components of the simple uniform structure. This is because as simple uniform structure for the complex uniform structure the results obtained by SCA1, SCA2, SCA5 and SCA6 are just the simple uniform structure. The results obtained by SCA3 are the same as those obtained by SCA4, and are much less accurate than those found by the other SCA methods. In contrast to the results obtained by SCA3 for simple uniform structure (Section 6.3.1), SCA3 did not get the same results as SCA1, SCA2, SCA5 and SCA6. This means the results obtained by SCA3 are only good for the 8 dimensional simple uniform structure. The transformation matrix for SCA1 is the same as that for SCA2 in each iteration, the transformation matrix for SCA3 is the same as that

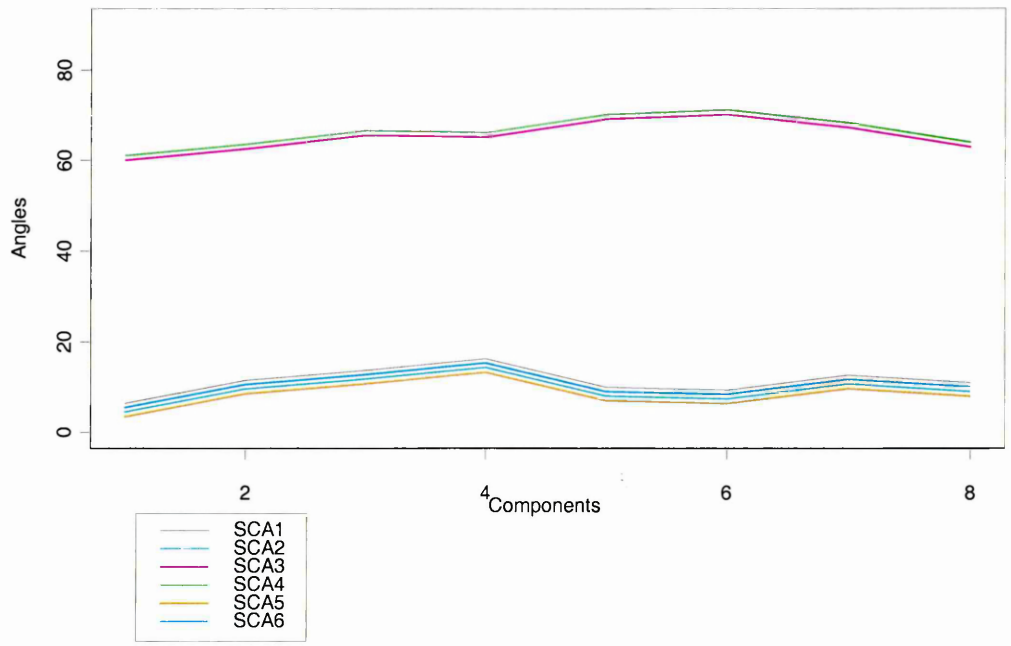

Figure 6.3: The accuracy of components produced by SCA methods for 8 dimensional complex uniform structure

for SCA4 in each iteration. That means for the complex uniform structure, the transformation matrices in each iteration are the same for single SCA methods with the same criterion whether the condition $l_r = l_s$ is imposed or not.

The results obtained by all the SCA methods are simple (Figure 6.4). The means of the components obtained by SCA3 and SCA4 are the same and are the most simple results. The components produced by other SCA methods have the same means, the means of the components produced by other SCA methods are just the means of the columns for the simple uniform structure. This is expected because the results obtained by SCA1, SCA2, SCA5 and SCA6 are just the simple uniform structure.

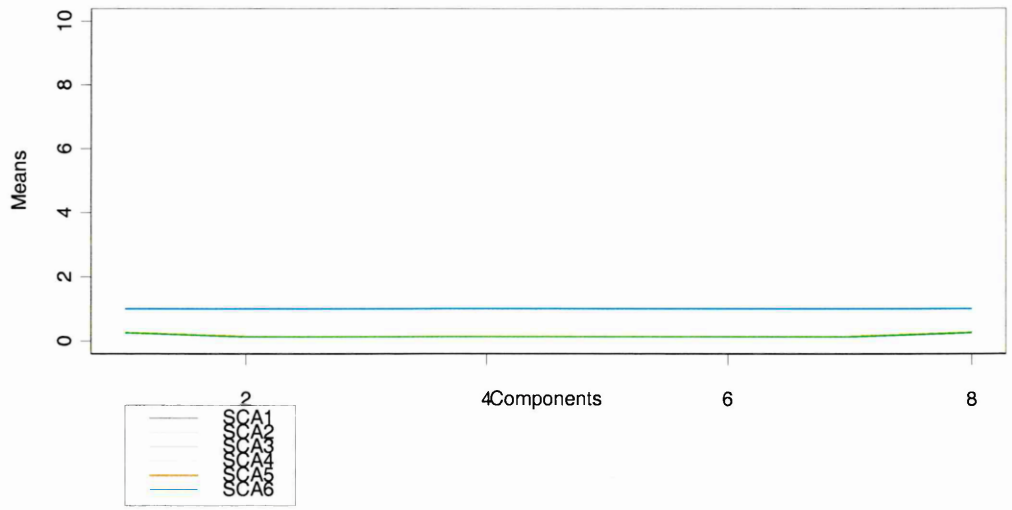

Figure 6.4: The simplicity of components produced by SCA methods for 8 dimensional complex uniform structure

6.4 Population results for 10 dimensional data, $k = 0$

Each SCA method will be applied to the variance-covariance matrices V_0 for 10 dimensional data in this section. Although the results obtained by SCA methods are different for simple structures and complex structures, the general conclusions about accuracy and simplicity of the components produced by SCA methods are the same as those for 6 and 8 dimensional data (the last paragraph of Section 4.7). So for 10 dimensional data, all the structures are considered together.

For 10 dimensional data, as for 6 dimensional data in Sections 4.5 and 4.6, generally SCA2, SCA5 and SCA6 are the most accurate methods. For 8 dimensional data, SCA1, SCA2, SCA5 and SCA6 get the most accurate results (Section 6.4). Generally, the results obtained by SCA methods with maximal improvement in variance are at least as accurate but not as simple as

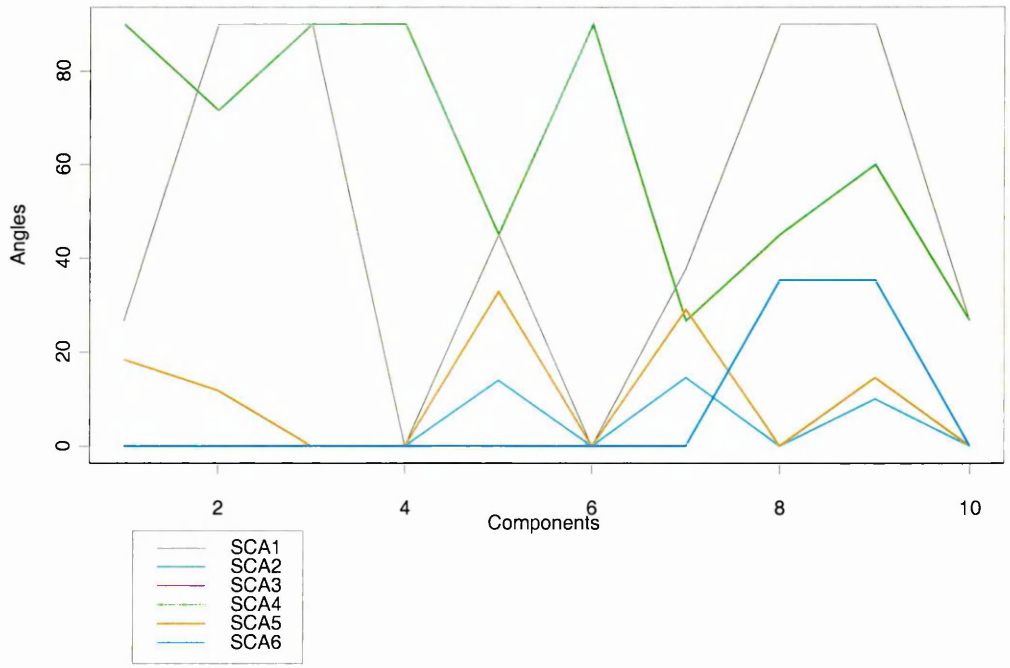

Figure 6.5: The accuracy of components produced by SCA methods for 10 dimensional simple uniform structure

those obtained by SCA methods with maximal variance. In other words, in general, the results obtained by SCA2 are more accurate but not as simple as those obtained by SCA4, the results obtained by SCA1 are more accurate but not as simple as those obtained by SCA3, the results obtained by SCA5 generally are more accurate but not as simple as those obtained by SCA6. The results obtained by the single SCA methods with maximal improvement in variance are at least as accurate as and more simple than those obtained by the multiple SCA method with maximal improvement in variance, i.e. generally the results obtained by SCA2 are at least as accurate as and more simple than those obtained by SCA5. The results obtained by the multiple SCA method with maximal variance are more accurate but not as simple as those obtained by single SCA method with maximal variance, i.e. the results obtained by SCA6 are more accurate but not as simple as those obtained by SCA4. The condition $l_r = l_s$ makes the results obtained

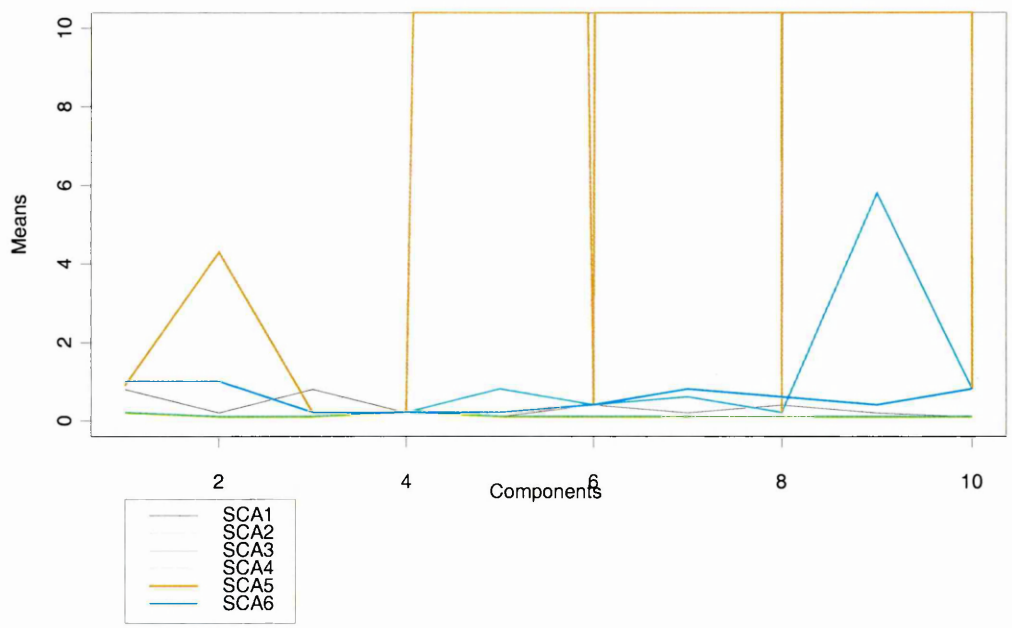

Figure 6.6: The simplicity of components produced by SCA methods for 10 dimensional simple uniform structure

by SCA1 and SCA3 no more accurate but more simple than those obtained by the corresponding SCA methods without this condition, i.e. the results of SCA1 are less accurate but more simple than those obtained by SCA2, the results of SCA3 are less accurate but more simple than those obtained by SCA4.

As a particular example, as for 8 dimensional data, consider the 10 dimensional simple uniform structure. SCA2 and SCA6 (Figure 6.5) get the first two components of the simple uniform structure back exactly, thus they are the best methods for the first 2 components for 10 dimensional simple uniform structure. Remember for 6 dimensional simple uniform structure, only SCA2 gets the simple uniform structure back (Figure 4.3). Comparing the results for the 10 dimensional simple uniform structure with those for the 8 dimensional simple uniform structure in Figure 6.1, SCA1, SCA3 and SCA5 are less accurate. I am not surprised that the results

for SCA1 and SCA3 are less accurate compared to the results obtained by SCA1 and SCA3 for 8 dimensional simple uniform structure because generally they are less accurate whatever the structure. However I am very surprised that the results of SCA5 are less accurate than those of SCA6. Remember that generally the results obtained by SCA5 are not less accurate than those of SCA6. But it is possible considering the simple components are obtained by the multiplication of the transformation matrices in each step. The more accurate results in each step does not always guarantee the more accurate results finally. All the results obtained by SCA methods are simple (Figure 6.6), i.e. all the means of the first two components produced by SCA methods are less than 10 for simple uniform structure. The means for the first two components obtained by SCA2 and SCA6 are the same as the actual means for the first two components of simple uniform structure because SCA2 and SCA6 retrieved the first two components of simple uniform structure. The first two components obtained by SCA1, SCA3 and SCA4 are more simple than the actual first two components for simple uniform structure. The mean of the first component obtained by SCA5 is slightly less than the actual mean, the mean of the second component obtained by SCA5 is more than the actual mean. But the difference is not very obvious. SCA2 and SCA6 are the best methods for 10 dimensional simple uniform structure.

6.5 Discussion and conclusion

In all the above examples of 8 and 10 dimensional data, in general, SCA2, SCA5 and SCA6 are the best methods for all the structures. This is the same as that for 6 dimensional data. As 6 dimensional data, the general conclusions about the accuracy and the simplicity of the components produced by SCA methods are the same whatever the dimension of the data (the last paragraph of Section 4.7).

As for 6 dimensional data, SCA methods can retrieve simple block and simple uniform structures (reason was given in Section 4.7). The best angles obtained by SCA methods for the complex structure generally are obtained if the SCA method can get the corresponding simple structure back exactly because the simple structure is generally the closest simple approximation of the corresponding complex structure.

SCA2 and SCA5 generally get the same results for the data introduced in Section 6.2. That means the multiplication of the transformation matrices by single updating with maximal improvement in variance (SCA2) are more likely to be the same as the multiplication of the transformation matrices by multiple updating with maximal improvement in variance (SCA5). The only difference is that SCA2 updates one pair and SCA5 updates as many pairs as possible at each iteration. Why SCA2 and SCA5 get the same results for most of the data introduced in Section 6.2 can be explained in a similar way like the example in Section 4.5.1 for 6 dimensional simple block structure or Section 6.3.1 for 8 dimensional simple uniform structure.

As the dimensionality increases, the difference between SCA5 and SCA6 reduces in each iteration, because SCA5 and SCA6 update as many pairs as possible, the larger the dimension of the data, the more pairs are updated generally in each iteration and all the pairs must be different in each iteration, this also leads to the transformation matrices for SCA5 and SCA6 in each iteration being more likely to be the same (example was given in Section 3.3.5). As said in Section 4.7 this process also reduces the difference between the maximal improvement in variance criterion and maximal variance criterion. So the difference between SCA5 and SCA6 reduce.

The results obtained by SCA1 for 8 dimensional data are very good, it seems that the condition $l_r = l_s$ does not prevent SCA1 from getting the correct results for 8 dimensional data, but the results obtained by SCA1 are not better than those obtained by SCA2. This is because for all the 8 dimensional data, in most of the iterations, SCA1 and SCA2 have the same transformation matrices. That is to say, for the 8 dimensional data in Section 6.2, generally the directions with maximal improvement in variance have the same length. In contrast the results obtained by SCA1 for 6 and 10 dimensional data are still disappointing. As for 6 dimensional data, SCA3 and SCA4 are the least accurate methods for large dimensional data, though the results obtained by SCA3 and SCA4 for the first two components are very similar. So generally SCA2, SCA5 and SCA6 are the best SCA methods whatever the structures and whatever the dimension of the data.

Chapter 7

Sample simulation results for 8 and 10 dimensional data

7.1 Introduction

In Chapter 6, the population results for 8 and 10 dimensional data in Section 6.2 was investigated. In this chapter, the sample simulation results for 8 and 10 dimensional data will be considered. The question that interests us here is whether the sample simulation results lead to the similar conclusions as those from the population results for large dimensional data. In particular, as the dimension increases, do SCA2, SCA5 and SCA6 tend to get the similar results? Are the sample simulation results obtained by SCA1 still good for 8 dimensional data?

As in Chapter 5, the sample simulation for large dimensional data is based on 500 samples each of 500 observations from a normal distribution $N(0, V_0)$. In Section 7.2, the sample simulation results produced by the SCA methods for 8 dimensional data will be discussed. In Section 7.3, the sample simulation results produced by the SCA methods for 10 dimensional data will be investigated. It is known that as the dimension increases, the number of the parameters to be estimated in a sample variance-covariance matrix increases so, in Section 7.4, the sample simulation results obtained by SCA methods will be investigated if the number of observations is taken to be 5000. The last section, Section 7.5, is the discussion and conclusion.

7.2 Sample simulation results for 8 dimensional data, $k = 0$

In this section, I am going to investigate the sample simulation results for 8 dimensional data when $k = 0$. As in Chapters 4 to 6, three types of structures are investigated, block structure, uniform structure and intermediate structure. Each structure has a simple form and a complex form. All the structures and eigenvalues here are the same as those introduced in Section 6.2.1. As in Chapter 5, in the following boxplots, three different colours are used. The black represents the difference between the median and the lower quartile, the blue represents the range between the median and the upper quartile. So that it is known exactly where is the median of the angles of each SCA method even if it equals one of the quartiles. The sample simulation results are displayed in the following order: the first two components produced by PCA, the first two components produced by SCA1 and so on to the first two components produced by SCA6. The boxplots represents the 500 angles between the original components and the simple components or the 500 means of absolute value of simple components for 8 dimensional data for the first two components of PCA and SCA (the definition of angles and means are given in Section 4.5). As in Chapter 6, first the sample simulation results for the 8 dimensional simple structures will be investigated.

7.2.1 Simple structures

For all the 8 dimensional simple structures, generally the results obtained by SCA2, SCA5 and SCA6 are the most accurate. The results obtained by all the SCA methods are simple. In spite of the results obtained by SCA methods for 8 dimensional data being different from those for the 6 dimensional data, the general conclusions about the accuracy and simplicity of components produced by SCA methods are the same as those for 6 dimensional data (the first paragraph of Section 4.7). Generally the results obtained by SCA methods with maximal improvement in variance are at least as accurate but not as simple as those obtained by SCA methods with maximal variance, i.e. the results obtained by SCA1 are more accurate but not

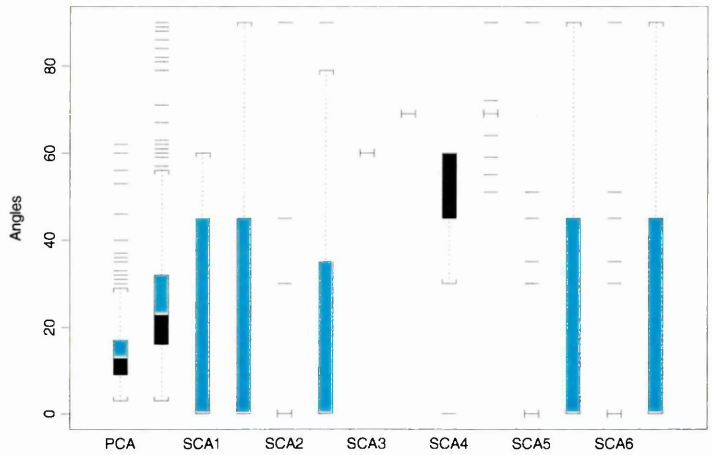

Figure 7.1: The accuracy of the first two components produced by PCA and SCA for 8 dimensional simple uniform structure

as simple as those obtained by SCA3, the results obtained by SCA2 are more accurate but not as simple as those obtained by SCA4, the results obtained by SCA5 and SCA6 are the same. The results obtained by single SCA method with maximal improvement in variance are at least as accurate as and more simple than those obtained by multiple SCA method with maximal improvement in variance (the results of SCA2 and SCA5 are the same for 8 dimensional data). The results obtained by multiple SCA method with maximal variance are more accurate but not as simple as the results obtained by single SCA method with maximal variance, i.e. the results obtained by SCA6 are more accurate but not as simple as those obtained by SCA4. The condition $l_r = l_s$ makes the results obtained by the SCA methods no more accurate but more simple than the results obtained by SCA methods without this condition (the results obtained by SCA1 and SCA2 are the same, generally the results obtained by SCA3 are less accurate but more simple than those obtained by SCA4). As with the population results, the results of SCA1 again are good both in accuracy and simplicity for the 8 dimensional simple structures. For the 8 dimensional simple block structure and uniform structures, the sample simulation results

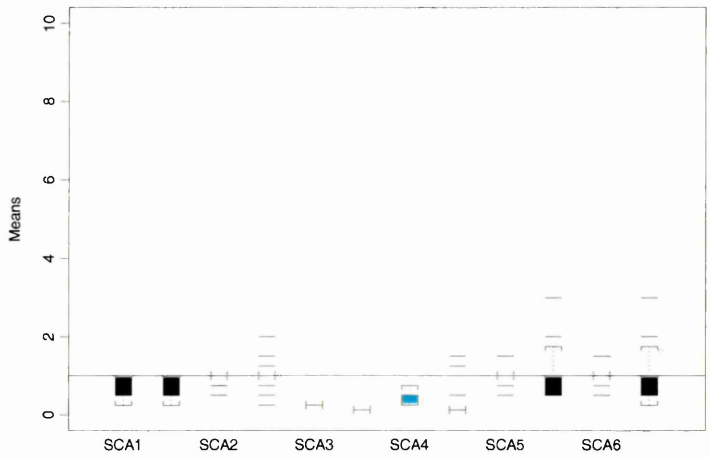

Figure 7.2: The means of the first two components produced by SCA for 8 dimensional simple uniform structure

obtained by SCA2, SCA5 and SCA6 are more accurate than those obtained by PCA. This is the same as the results for 6 dimensional data. For the 8 dimensional simple intermediate structure, the sample simulation results obtained by PCA are slightly more accurate than those obtained by SCA2, SCA5 and SCA6.

As a particular example, as for the population results, consider the sample simulation results for the 8 dimensional simple uniform structure. Figure 7.1 shows that, most of the time SCA2, SCA5 and SCA6 retrieved the simple uniform structure exactly. In fact, SCA5 and SCA6 get the same results every time. Overall the results obtained by SCA1, SCA5 and SCA6 for 8 dimensional data are more accurate than those for 6 dimensional data. For example, consider the results for 8 dimensional simple uniform structure in Figure 7.1 with the results for 6 dimensional simple uniform structure in Figure 5.3. I am not surprised about this because for the 8 dimensional simple uniform structure, the population results of SCA1, SCA5 and SCA6 are more accurate than the population results for 6 dimensional data (Figures 4.3 and 6.1). Generally, the results obtained by SCA1, SCA2, SCA5 and SCA6 are more accurate

than those obtained by PCA. Unfortunately, the results obtained by SCA3 and SCA4 are still very disappointing (Sections 4.2, 5.2 and 6.3.1) because of the large angles for the first two components. For example, the results obtained by SCA3 and SCA4 in Figures 7.1 and 5.3 are still very disappointing. I am very surprised about the bad results obtained by SCA3 for 8 dimensional simple uniform structure (Figure 7.1) considering the population results for 8 dimensional simple uniform structure (Figure 6.1). This indicates that only in the ideal case (the population results) are the results obtained by SCA3 for 8 dimensional simple uniform structure more accurate than those obtained by SCA4. So generally, the results obtained by SCA3 are less accurate than those of SCA4.

In contrast to SCA1, SCA5 and SCA6, the results obtained by PCA for 8 dimensional data are less accurate than those for the corresponding simple structures for 6 dimensional data. For example, for the simple uniform structure, you can see the results obtained by PCA in Figure 7.1 are less accurate than those obtained by PCA in Figure 5.3. This is expected, because as the dimension increases, the number of the parameters to be estimated in a sample variance-covariance matrix increases greatly. For 6 dimensional data, the number of parameters to be estimated is 21, for 8 dimensional, the number of parameters to be estimated is 36. However, the same number of observations, 500, is used in each data set.

The results obtained by all the SCA methods are simple (Figure 7.2) for the 8 dimensional simple uniform structure. The actual means of the first two components are 1, so a line of $y = 1$ is plotted on the boxplot of the first two components of SCA methods (Figure 7.2). It is clear that almost all the means for the first two components obtained by SCA3 and SCA4 are less than the means of the actual first two components. The medians of the means of SCA1 are the same as the actual means for the first two components. Almost all the means for the first two components obtained by SCA2 are the same as the actual mean. Almost all the means obtained by SCA5 and SCA6 are the same as the actual mean for the first component, the medians of the means for the second component obtained by SCA5 and SCA6 are the same as the actual mean, this can be expected because SCA1, SCA2 almost every time retrieve the first two components of the simple uniform structure, SCA5 and SCA6 most of the times retrieve the first two components of simple uniform structure.

7.2.2 Complex structures

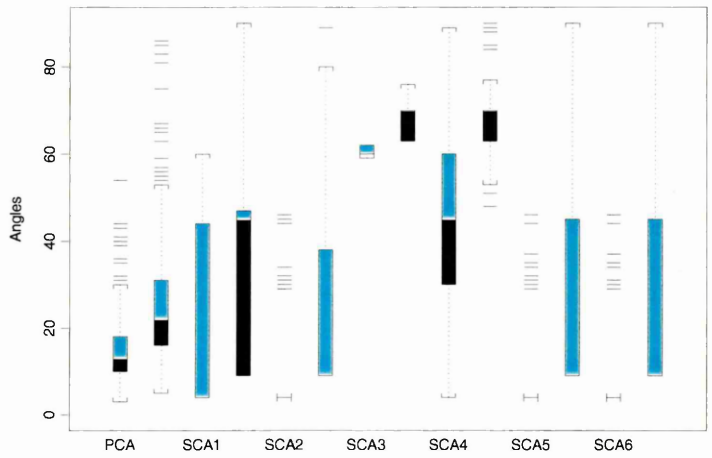

Figure 7.3: The accuracy of the first two components produced by PCA and SCA for 8 dimensional complex uniform structure

For 8 dimensional complex structures, again generally, the results obtained by SCA2, SCA5 and SCA6 are the most accurate. In contrast to the population results, SCA1 is not so good for the 8 dimensional complex structures. The results obtained by SCA3 and SCA4 are still very disappointing (Sections 5.2, 6.3.1 and 7.2.1) because of the large angles for the first two components. The general conclusions about the accuracy and simplicity of the components produced by SCA methods are the same as those for the 6 dimensional data and the 8 dimensional simple structures given in the last paragraph of Section 4.7 and Section 7.2.1. In contrast to the results for the 6 dimensional data, for the 8 dimensional complex block and uniform structures, the sample simulation results obtained by SCA2, SCA5 and SCA6 are more accurate than those obtained by PCA. As for the 6 dimensional data, for the 8 dimensional complex intermediate structure, the results obtained by SCA2, SCA5 and SCA6 are less accurate than those obtained by PCA. As in Section 6.3.2, details are given below only for the 8 dimensional complex uniform structure.

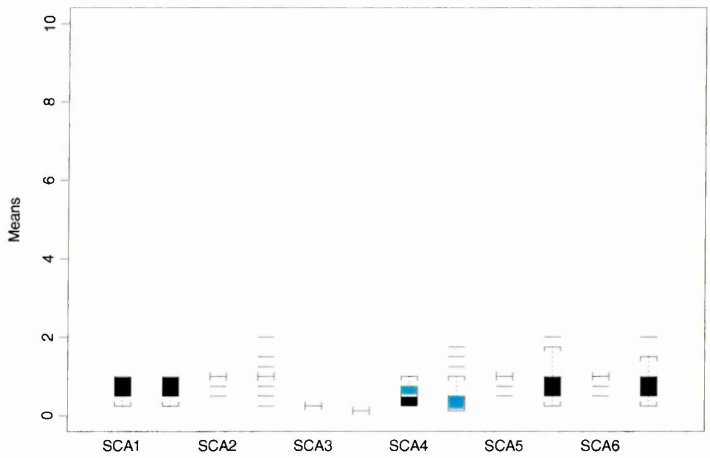

Figure 7.4: The means of the first two components produced by SCA for 8 dimensional complex uniform structure

The results for the first two components of the complex uniform structure obtained by SCA2, SCA5 and SCA6 are the most accurate (Figure 7.3). Only a few times did SCA2, SCA5 and SCA6 not retrieve the first component of simple uniform structure exactly. Also most of the time SCA2, SCA5 and SCA6 retrieved the second component of the simple uniform structure exactly. That is why the medians of the angles of SCA2, SCA5 and SCA6 are just the angles between the first two components of the simple uniform structure and the first two components of the complex uniform structure. As for 8 dimensional simple structures and 6 dimensional data, the results obtained by SCA3 and SCA4 are disappointing. In contrast to 8 dimensional simple uniform structure (Figure 7.1), the results of SCA1 are not so accurate as those of SCA2. For 8 dimensional complex structures, the sample simulation results of SCA1 are much less accurate than those obtained by SCA2. This is different from the population results (Figure 6.3) for 8 dimensional complex uniform structure where SCA1 and SCA2 get the same results. This means for the sample simulation results for 8 dimensional complex uniform structure, the transformation matrix of SCA1 is different from that of SCA2 in each iteration. Comparing

the results in Figures 5.9 and 7.3, the results of SCA2, SCA5 and SCA6 for 8 dimensional complex uniform structure are more accurate than those for 6 dimensional complex uniform structure. This is expected. For the population results, the components obtained by SCA2, SCA5 and SCA6 for 8 dimensional complex uniform structure are the 8 dimensional simple uniform structure. For the population results for 6 dimensional complex uniform structure, only SCA5 gets the first two components of the 6 dimensional simple uniform structure.

Again, the results obtained by PCA for complex structure are less accurate than those of the corresponding complex structures for 6 dimensional data (Section 5.2). This is expected, because as the dimension increases, the number of the parameters to be estimated in a sample variance-covariance matrix increases greatly, but the the same number of observations is used, 500, in each dataset in 8 dimensional data as for 6 dimensional data.

The results obtained by all the SCA methods (Figure 7.4) are simple for the first two components for the complex uniform structure. All the conclusions about the accuracy and simplicity of the results are the same as those for 6 dimensional data and 8 dimensional simple structures given in Section 4.7.

7.3 Sample simulation results based on 10 dimensional data, $k = 0$

In this section, the sample simulation results for all the 10 dimensional data are investigated together. All the structures and eigenvalues here are the same as those in Section 6.2.1. As in previous sections, I will investigate the sample simulation results for three types of structures, block structure, uniform structure and intermediate structure.

For 10 dimensional data, in general, the results obtained by SCA2 are the most accurate. As for 8 dimensional data, for both simple and complex block and uniform structures, the results obtained by SCA2 are slightly more accurate than those obtained by PCA. For 10 dimensional simple and complex intermediate structures, the results obtained by PCA are slightly more accurate than the best results obtained by SCA methods. Furthermore generally the results

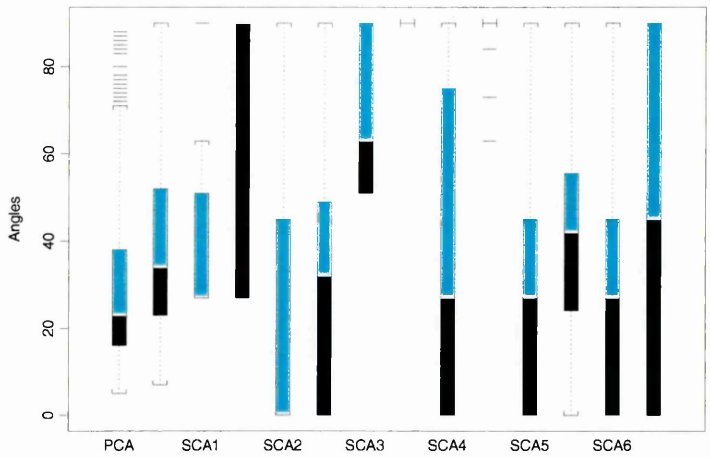

Figure 7.5: The accuracy of the first two components produced by PCA and SCA for 10 dimensional simple uniform structure

obtained by SCA2 are simple, i.e. generally the components produced by SCA2 are small-valued integers. So as for 8 dimensional data, the results obtained by SCA2 are better than those obtained by PCA for block and uniform structure. As for the population results for 10 dimensional data in Section 6.4, again the results obtained by SCA1, SCA3 and SCA4 generally are disappointing because of the large angles for the first two components. The general conclusions about the accuracy and simplicity of the components produced by SCA methods are the same as those for the 6 dimensional data given in Section 4.7 and the 8 dimensional data given in Section 7.2.1. The results obtained by PCA for 10 dimensional data are less accurate than those for 6 and 8 dimensional data. I am not surprised about this, for 10 dimensional data, there are more parameters (55 parameters) to be estimated in the sample variance-covariance matrix than those for 6 and 8 dimensional data (21 and 36 parameters to be estimated respectively), so generally more observations are needed for 10 dimensional data. But the same number of observations are made as for 6 and 8 dimensional simulations.

As a particular example, consider the sample simulation results for 10 dimensional simple

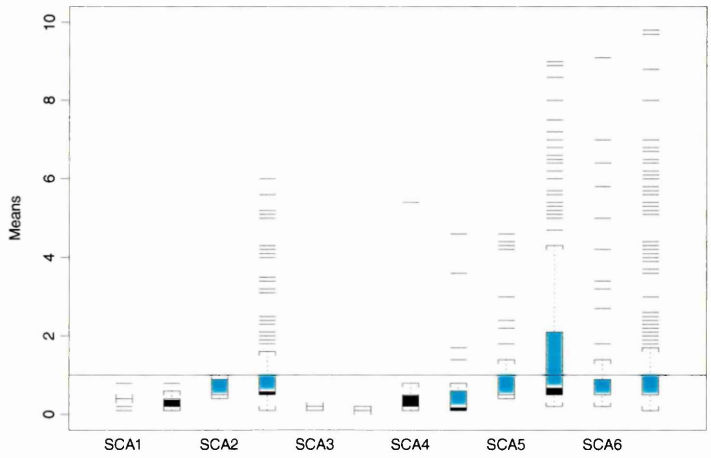

Figure 7.6: The means of the first two components produced by SCA for 10 dimensional simple uniform structure

uniform structure (Figures 7.5, 7.6). Most of the results obtained by all the SCA methods are simple (Figure 7.6). So I just pay more attention on the accuracy (Figure 7.5) of the sample simulation results. Also I am going to focus on SCA2, SCA5 and SCA6 because they are the most accurate SCA methods generally. In contrast to the corresponding population results, only the median of the angles of SCA2 for the first component is 0, medians of the other angles obtained by SCA2, SCA5 and SCA6 are not 0. Checking the results obtained by SCA2, SCA5 and SCA6, about 50% of the first components produced by SCA2 are the original first component. About 45% of the first components produced by SCA6 are the original first component. About 42% of the second components produced by SCA2 and SCA6 are the original second component. However only 26% of the first components produced by SCA5 are the original first component, about 9% of the second components produced by SCA5 are the original second component. The big difference between the sample simulation results obtained by SCA5 and SCA6 for 10 dimensional simple uniform structure is not surprising considering the corresponding population results. In the ideal case, the population results, the results obtained by SCA6 are much more

accurate than the results obtained by SCA5 (Figure 6.5).

All the general conclusions about the accuracy and simplicity of the results are the same as those for 6 and 8 dimensional data given in Sections 4.7 and 7.2.1.

7.4 Sample simulation results with 5000 observations

It is known that more parameter have to be estimated in a sample variance-covariance matrix as the dimension of the data increases. In order to get more accurate results, so sample simulation results obtained by SCA methods in this section are based on data sets of 5000 observations.

In general, the sample simulation results for large dimensional data using 5000 observations are much more accurate than those using 500 observations. However the general conclusions about the accuracy and simplicity of the components produced by SCA methods are the same as before (Section 4.7). So I only give details for 8 and 10 dimensional simple uniform structures.

7.4.1 The sample simulation results for 8 dimensional simple uniform structure

For 8 dimensional simple uniform structure, comparing the results in Figures 7.7 and 7.8 with those in Figures 7.1 and 7.2, it is clear that the results obtained by SCA methods using 5000 observations are much more accurate and more simple than those using 500 observations. Recall from Section 6.3.1 that for the population results, SCA1, SCA2, SCA3, SCA5 and SCA6 retrieved the first two components of 8 dimensional simple uniform structure. Here with data sets of 5000 observations are used almost every time SCA1, SCA2, SCA5 and SCA6 retrieved the 8 dimensional simple uniform structure exactly. The only difference between the population results and sample simulation results is that the sample simulation results obtained by SCA3 are disappointing. This indicates that SCA3 only gets the best results in the ideal population situation. The results obtained by SCA1, SCA2, SCA5 and SCA6 are much more accurate than those obtained by PCA. Also the results obtained by SCA1, SCA2, SCA5 and SCA6 must be simple because almost every time they retrieved the first two components of 8 dimensional sim-

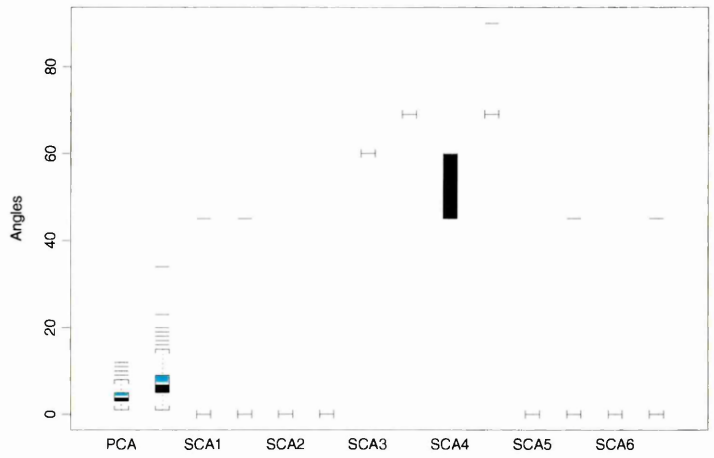

Figure 7.7: The accuracy of the first two components produced by PCA and SCA with 5000 observation for 8 dimensional simple uniform structure

ple uniform structure. So the results obtained by SCA1, SCA2, SCA5 and SCA6 are better than those obtained by PCA. Comparing the results in Figure 7.7 with those in Figure 5.3, the results obtained by SCA2 are the same as those for 6 dimensional data. Also SCA1, SCA5 and SCA6 almost every time get the correct answer. So the results obtained by SCA1, SCA5 and SCA6 (Figure 7.7) are much more accurate than those of SCA1, SCA5 and SCA6 for 6 dimensional data (Figure 5.3). This is expected comparing the population results for 6 dimensional data (Figure 4.3) and with the results for 8 dimensional data (Figure 6.1).

7.4.2 The sample simulation results for 10 dimensional simple uniform structure

For 10 dimensional simple uniform structure, SCA2 and SCA6 just a few times did not retrieve the first two components of the simple uniform structure (Figure 7.9). This is expected considering the population results for 10 dimensional simple uniform structure, where SCA2 and SCA6 retrieve the first two components of the 10 dimensional simple uniform structure exactly.

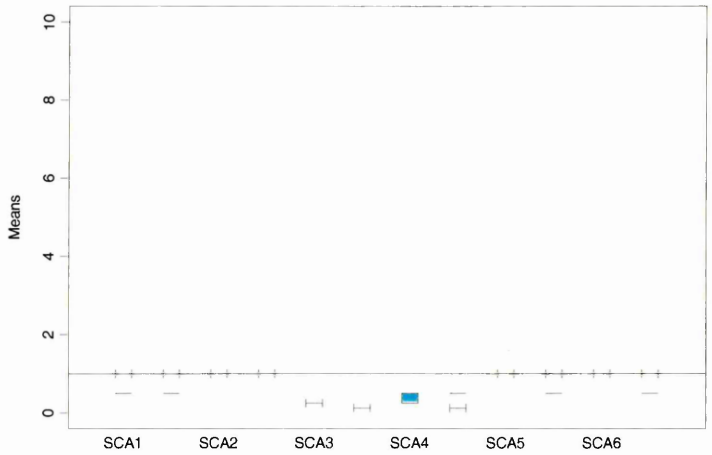

Figure 7.8: The means of the first two components produced by SCA with 5000 observation for 8 dimensional simple uniform structure

Also this is expected considering the sample simulation results with 5000 observations for 8 dimensional data, a particular example is the 8 dimensional simple uniform structure (Figure 7.7). This is because for 8 dimensional data if an individual SCA method can retrieve a simple structure exactly for the population results, this SCA method almost every time retrieves this simple structure for the sample simulation results with 5000 observations. I am not surprised about the the big difference between the results of SCA5 and SCA6 just comparing the population results for 10 dimensional simple uniform structure in Figure 6.5 with the sample simulation results for this structure in Figure 7.5. Actually, about 50% of the first two components produced by SCA5 are the same as the original first two components. This just indicates that compared to SCA5, SCA2 and SCA6 are more likely to retrieve the first two components of 10 dimensional simple uniform structure. Again, as the results for 10 dimensional simple structure with 500 observations (Figure 7.5), the results obtained by SCA1, SCA3 and SCA4 are disappointing.

All the results produced by SCA methods are very simple (Figure 7.10). The components obtained by SCA1, SCA3 and SCA4 are more simple than the actual components. Almost all

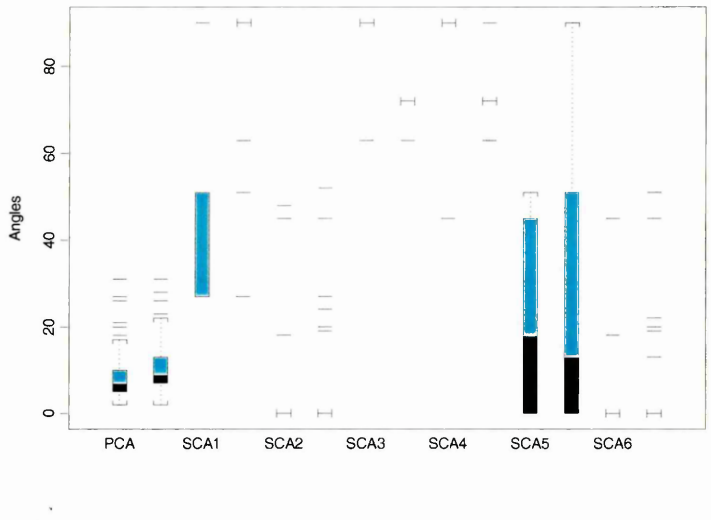

Figure 7.9: The accuracy of the first two components produced by PCA and SCA with 5000 observation for 10 dimensional simple uniform structure

the means obtained by SCA2 and SCA6 are the same as the actual means because SCA2 and SCA6 almost every time get the first two components back exactly.

7.5 Discussion and conclusion

In general, SCA2, SCA5 and SCA6 are the best methods for all the large dimensional data structures. All in all, whatever the dimension of the data, generally SCA2, SCA5 and SCA6 are the best SCA methods, the results obtained by SCA1, SCA3 and SCA4 are not so good as the results obtained by SCA2, SCA5 and SCA6. The general conclusions about the accuracy and simplicity of the results obtained by SCA methods are the same as those for 6 dimensional data given in the last paragraph of Section 4.7. For large dimensional data, as for 6 dimensional data, the same conclusions are obtained by the population results and sample simulation results. So it is easy to check which SCA method is the most accurate or the most simple by just considering the population results.

If the number of observations in a data set was increased, it was shown that the sample

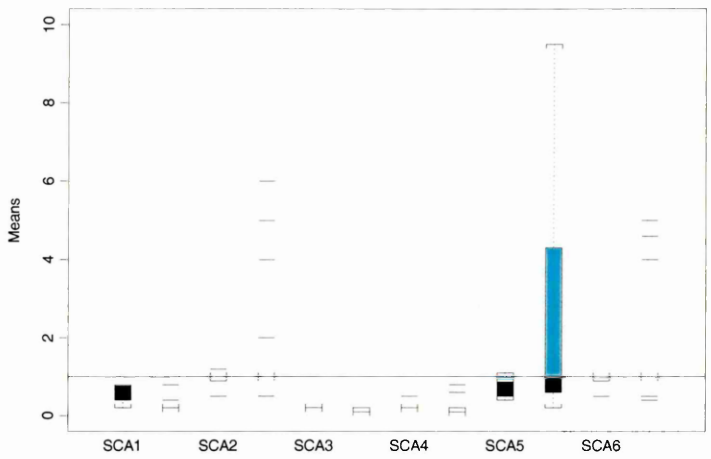

Figure 7.10: The means of the first two components produced by SCA with 5000 observation for 10 dimensional simple uniform structure

simulation results obtained by SCA methods improved greatly. This is to be expected because as the size of the sample increases, the estimated parameters are more accurate, i.e. the elements of the variance-covariance matrix of the sample are closer to the elements of the variance covariance matrix V_0. So the variance-covariance matrix of the sample is closer to the variance-covariance matrix V_0 when the sample size is increased.

In contrast to the sample simulation results for the 6 dimensional data, for large dimensional block and uniform structures (both simple and complex), the most accurate sample simulation results obtained by SCA method were generally more accurate than the sample simulation results obtained by PCA. The reason for SCA methods to get more accurate sample simulation results than those obtained by PCA for simple block and simple uniform structures have been given in Section 5.5. For large dimensional complex block and complex uniform structures, this is true because the angles between the simple structure and corresponding complex structure are smaller than the difference between the sample simulation results of PCA and SCA for simple block and simple uniform structures. Recall that SCA2, SCA5 and SCA6 get the simple block structure

for both simple and complex block structures, they also get the simple uniform structure for both simple and complex uniform structures. So the most accurate sample simulation results of SCA for complex block and complex uniform structures are the angles between simple structure and corresponding complex structure. The sample simulation results of PCA generally are no more accurate than 8° and 11° for the first two components respectively for block and uniform structures given in Section 6.2 whatever the structure of the data. The sample simulation results of SCA for the first two components of simple block and simple uniform structures generally are 0° and 0° respectively, whereas the angles between the first two components of simple block and complex block structures are 4° and 5° respectively, the angles between the first two components of the simple uniform and complex uniform structures are 4° and 10° respectively. PCA has no bias but large variance, whereas the angles obtained by SCA has a discrete distribution, SCA has some bias but small variance for complex block and complex uniform structures, so for the complex block and complex uniform structures, the sample simulation results obtained by SCA are more accurate than those obtained by PCA.

SCA methods are not good for intermediate structure as shown by the bigger angles for all the examples. One reason is that $k = 0$ may not be a good choice for this structure because the loadings in the transformation matrix at each iteration are not similar to the loadings of simple intermediate structure when $k = 0$ for two dimensional data. So SCA methods when $k = 0$, 1 and 2 will be investigated in Chapter 8. Only the population results are investigated because generally the population results and the sample simulation results gave the same conclusions.

Chapter 8

Further population results

8.1 Introduction

In Chapters 4 to 7, only the results obtained by SCA methods when $k = 0$ were investigated. As k increases, what happens to the results obtained by SCA methods? Also if the same eigenvector matrices in Sections 4.3 and 6.2 are used, but with different eigenvalues, do the results produced by SCA methods change? I will try to answer these two questions in this chapter.

It was shown in Chapter 3 that as k increases, the number of directions available for the simplicity preserving transformations at the heart of SCA methods increases. Also the directions for $k + 1$ include all the directions for k. Also in Chapter 3, it was proved that the bigger the maximal variance obtained by single SCA methods for any given directions r and s at a given step, the more accurate results obtained by single SCA methods for any given directions r and s in this step. So I suppose the results of SCA methods with maximal variance generally become more accurate as k increases. It was also shown in Chapters 4 to 7 that the results of SCA methods with maximal improvement in variance generally are more accurate than the results of SCA methods with maximal variance. So I expect that SCA methods with maximal improvement in variance should also get more accurate results as k increases generally. In other words, all the SCA methods are expected to obtain more accurate as k increases.

In this chapter, I only consider the population results obtained by SCA methods because in

Chapters 4 to 7, the population results and sample simulation results produced by SCA methods give the same conclusions whatever the structures and the dimension of the data. In Section 8.2, the population results when $k = 0$, 1 and 2 will be compared to check if the expectation that as k increase is correct. In Section 8.3, the influence of the eigenvalues on the results of the first two components will be discussed. Three different vectors of eigenvalues will be used for each data set in this chapter. One vector is the original vector of eigenvalues, i.e. the eigenvalues used in Chapters 4 and 5 or 6 and 7, where all the eigenvalues are different. The second vector of eigenvalues will have an extremely big difference between the first two elements, but the other elements are the same as the original eigenvalues. So under this second case, the first component is very important. The third vector will have an extremely small difference between its first two elements but the first two values are much bigger than the rest, the other elements are the same as the original eigenvalues. So in this third case, the first two components are very important. Section 8.4 briefly discusses the effect of k and of different vectors of eigenvalues for 8 and 10 dimensional data. The last section, Section 8.5, contains the discussion and conclusion.

The results obtained by different SCA methods using different k and different pattern of eigenvalues have been investigated. From the results obtained by different SCA methods in Chapters 4 to 7, it is shown that SCA2 is one of the best method for all the structures. So only the population results obtained by SCA2 are extracted. As in Chapters 4 to 7, only the first two components are investigated in this chapter.

8.2 Results for different k for 6 dimensional data

In this section, the population results obtained by SCA2 when $k = 0$, 1 and 2 will be compared for the 6 dimensional simple and complex block structures, uniform structures and intermediate structures.

8.2.1 Simple structures

Consider the population results for the 6 dimensional simple structures. When $k = 0$ SCA2 retrieves the simple block and simple uniform structures. So the results obtained by SCA2 for

the simple block and simple uniform structures using other k cannot do better than these results for the first two components because these results are the same as the first two components of the simple structure. The good news is that the first two components of simple intermediate structure are retrieved by SCA2 when $k = 1$. The results obtained by other SCA methods with $k = 0$ and higher k generally are worse than the results obtained by SCA2 when $k = 1$ for simple intermediate structure. Generally the smaller k, the more simple the components produced by SCA2. The results obtained by SCA2 applied to larger k generally are neither always more accurate nor as simple as those applied to smaller k.

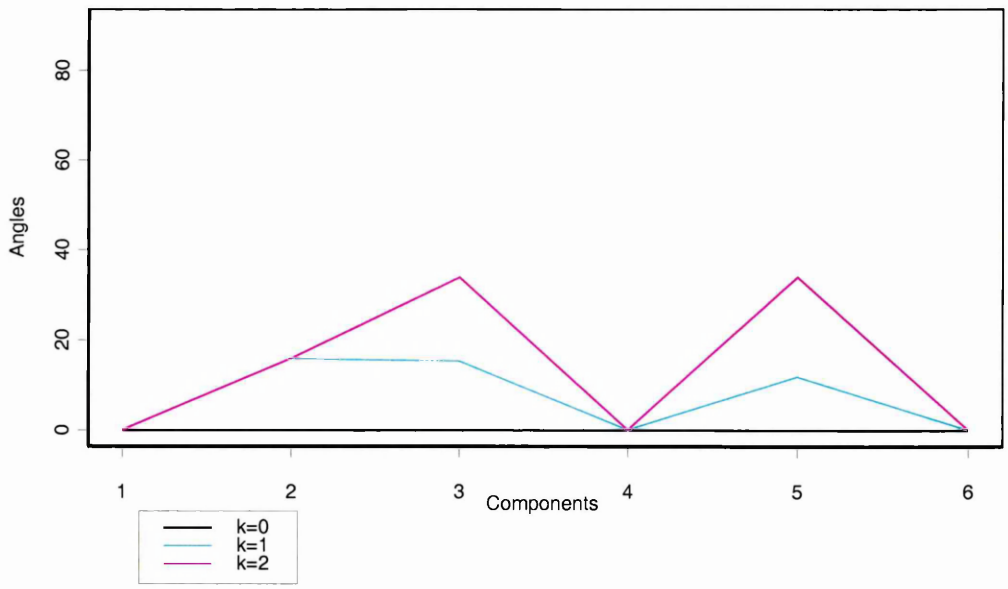

Figure 8.1: The accuracy of components produced by SCA2 when k is different for 6 dimensional simple block structure

I give details for the simple block structure; other simple structures were investigated in a similar way. First, consider the accuracy of the components. When $k = 0$, SCA2 gets the most accurate results for the simple block structure because the original components are retrieved

$k = 0$ are more simple than those when $k = 1$ and 2. Why the components obtained by SCA2 with smaller k generally are more simple than those with bigger k? There are two reasons. The first reason is whatever the structure of the data as k increases, the steps taken by SCA2 generally increase: more steps taken generally means less simple components produced. For example, for simple block structure, SCA2 took 5 steps when $k = 0$ and 1 and took 11 steps when $k = 2$. The other reason is the bigger k, generally the longer the length of the updating directions. As shown in Section 3.3.2, the transformation matrix for two-dimensional data at each step when $|b| < 1$ is $\begin{pmatrix} 2^k & il_2^2 \\ i & -2^k l_1^2 \end{pmatrix}$, when $|b| \geq 1$, $\begin{pmatrix} i & 2^k l_2^2 \\ 2^k & -il_1^2 \end{pmatrix}$. where $b = \frac{i}{2^k}$ or $b = \frac{2^k}{i}$, $i = -2^k, -2^k + 1, \ldots, 2^k$. The larger k, the larger the elements in the transformation matrix generally, so the larger k, generally the longer the lengths of the updated two directions. For more than two dimensional data using SCA2, the updating two columns (directions) at each step are as in the above matrices, the other directions are the same as the directions in the previous step. Given the transformation matrix mentioned before and the steps SCA2 with different k taken, the smaller the k, generally the more simple the results are obtained by SCA2.

8.2.2 Complex structures

In this subsection, I am going to consider the population results for the 6 dimensional complex structures. For the complex block structure, the best results are obtained when $k = 0$ and 1. For the complex uniform and complex intermediate structures, the best results are obtained when $k = 1$. As with simple structures, as k increases, the number of steps taken by SCA2 generally increases. Generally the smaller k, the more simple the components produced by SCA2. Again, the results of SCA2 using a larger k were generally neither always more accurate nor as simple as those using a smaller k.

As for simple structures, only more details of complex block structure are given. First consider the accuracy of components produced by SCA2 using different values of k. When $k = 0$ and 1 SCA2 gets the first two components of the corresponding simple block structure back exactly (Figure 8.3). So these results are just the angles between the first two components of simple block structure and complex block structure.

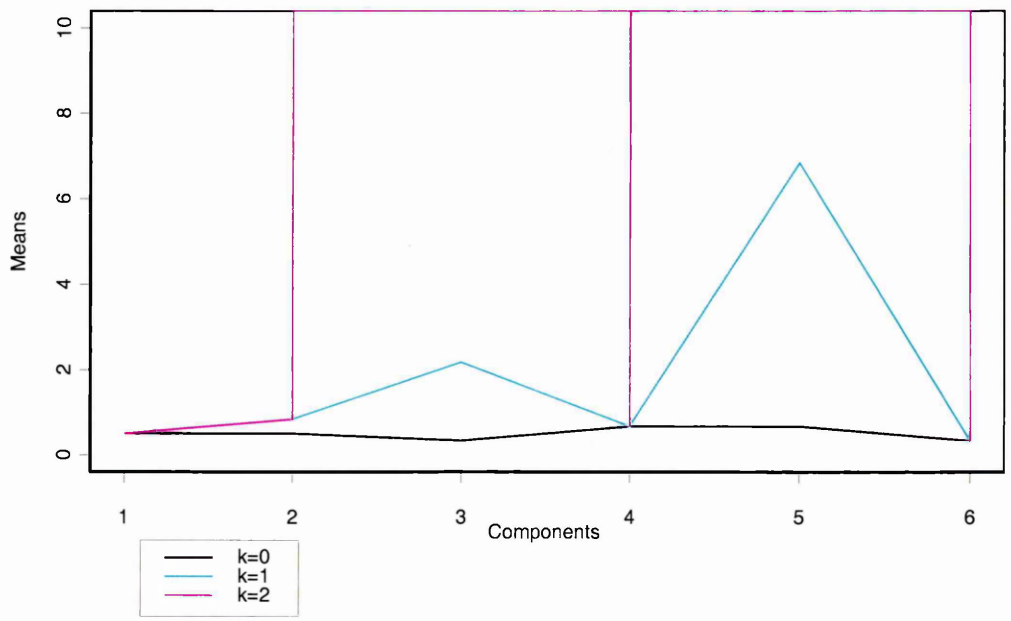

Figure 8.2: The simplicity of components produced by SCA2 when k is different for 6 dimensional simple block structure

exactly (Figure 8.1). The results obtained by SCA2 when $k = 1$ and 2 are worse than those obtained by SCA2 when $k = 0$. However given that at every step more directions for the simplicity preserving transformation were available with $k = 2$ compared with $k = 0$ and 1, it is surprising that the most accurate results are not obtained for all the simple structures using $k = 2$ because I suppose the higher the k the more accurate the results generally. In fact, out of the chosen k the results using $k = 2$ are proved to be the least accurate for the simple block and the simple intermediate structures. For the simple uniform structure, the results obtained by SCA2 when $k = 2$ are the second accurate.

So for the simple block structure, I will compare the transformation matrix of SCA2 when $k = 0$ and $k = 2$ at each iteration to see why SCA2 get the best results when $k = 0$ and SCA2 get the worst results when $k = 2$. For convenience, only comparing the first three steps taken

by SCA2 when $k = 0$ and 2. In the first two steps SCA2 updates the same pair of directions in the same way. So the transformation matrices are the same in the first two steps. In the third step, the same pair of directions are updated, i.e. directions d_1 and d_3, but the directions with maximal improvement in variance at this step are different. In other words, for different value of k the updates are on the same pair of directions but in different ways. This is because when $k = 2$, more directions are available in the two dimensional subspace including directions d_1 and d_3. So at this step, the best direction that leads to a maximal improvement in variance (the maximal improvement in variance is 1.2 when $b = 2$) is not available when $k = 0$ (the maximal improvement in variance is 1.0 when $b = 1$). This means that in the third step, the maximal improvement in variance obtained by SCA2 when $k = 2$ is more than that when $k = 0$. However if the maximal iterations are chosen as 3, the results when $k = 2$ are less accurate than those when $k = 0$. In this case, the angles obtained by SCA2 using $k = 2$ is 45.0°, 56.8°, 45.0°, 90.0°, 60.0° and 90° respectively, the angles obtained by SCA2 using $k = 0$ is 45.0°, 54.7°, 45.0°, 30.0°, 30.0° and 30° respectively. That is to say even though, at each step, the maximal improvement in variance when $k = 2$ is at least as big as that when $k = 0$ for the first three steps, this does not guarantee the more accurate results overall being obtained when $k = 2$. Moreover, the maximal improvement in variance obtained by SCA2 with larger k at each step is not always more than the maximal improvement in variance obtained by SCA2 with a smaller k because from some step the new variance-covariance matrices are different for different k, it certainly should be true for a given step if the same variance-covariance matrix is used for different k at this step. For example, for simple block structure, from step 4, the corresponding variance-covariance matrix are different at the same step for $k = 0$ and 2 and SCA2 took more steps when $k = 2$. Once the $k = 0$ and $k = 2$ steps have diverged, then $k = 2$ may be on a route that does not allow it to get back to the simplest solution found by $k = 0$. That is why it is possible for the results obtained by SCA2 when $k = 2$ to be less accurate than those when $k = 0$.

Next, consider the simplicity of the components produced by different values of k for the simple structures. As before only give more details for simple block structure. For the first two components of simple block structure, the means of the first two components obtained by SCA2 when $k = 1$ and 2 are the same (Figure 8.2). The components produced by SCA2 when

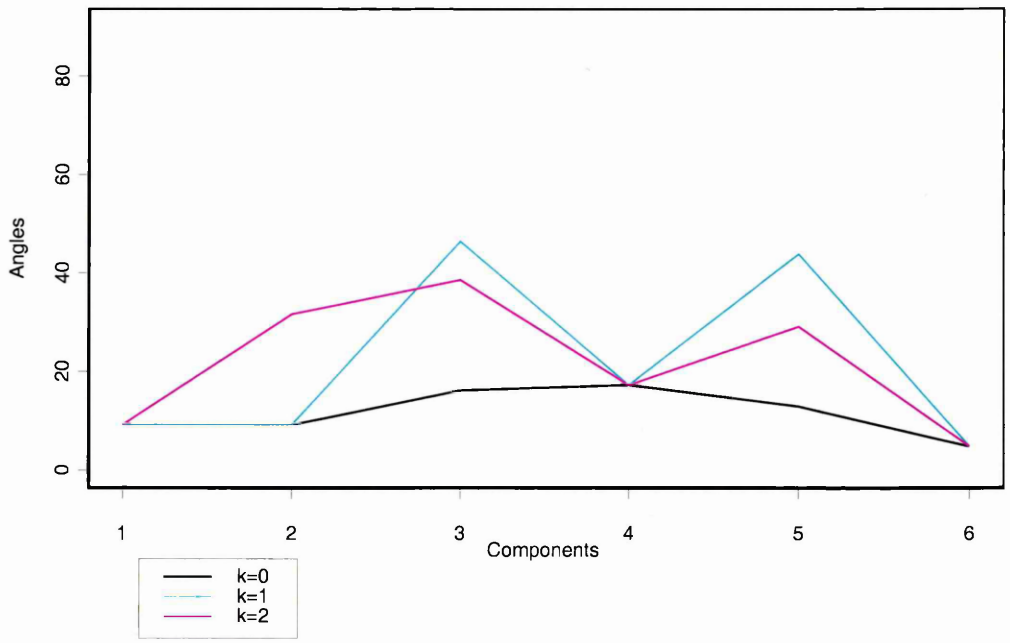

Figure 8.3: The accuracy of components produced by SCA2 when k is different for 6 dimensional complex block structure

Next consider the simplicity of components produced by SCA2 using different values of k. For complex block structure, when $k = 0$ and 1 SCA2 took 5 steps, when $k = 2$ SCA2 took 11 steps. As k increases, the steps taken by SCA2 generally increases. This is the same as what had been obtained for simple structures. The first two components obtained by SCA2 using different k are simple (Figure 8.4). The components produced by SCA2 when $k = 0$ and 1 are more simple than those produced by SCA2 when $k = 2$.

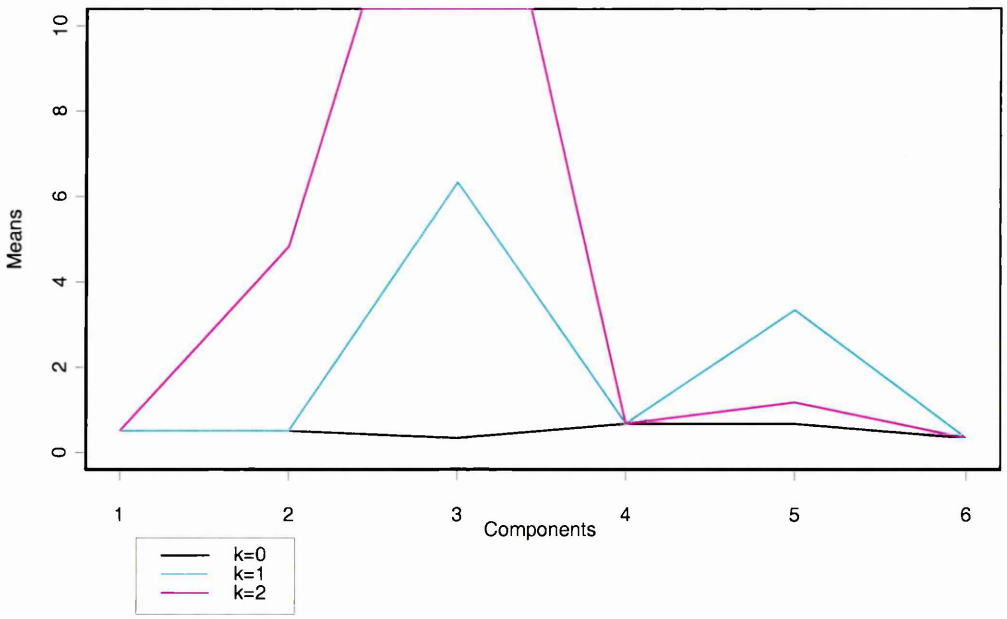

Figure 8.4: The simplicity of components produced by SCA2 when k is different for 6 dimensional complex block structure

8.3 Results based on different eigenvalues for 6 dimensional data, $k = 0$

In Chapters 4 and 5, the same eigenvalues $(12, 8, 6, 4, 2, 1)$ (original eigenvalues in Section 4.3) were used. In this section, two situations that may arise in theory will be discussed. One is equality of eigenvalues (suppose there are q equal eigenvalues), another is a big difference between eigenvalues. For convenience, in this section, the cases when the first two eigenvalues have big difference or they are very similar are considered. So two new sets of eigenvalues $(10000, 1000, 6, 4, 2, 1)$ and $(101, 100, 6, 4, 2, 1)$ but the eigenvector matrices are the same as the structures in Section 4.3 are used. The results obtained by SCA2 using three sets of eigenvalues (original eigenvalues and two new sets of eigenvalues) are discussed. The vectors of the three

eigenvalues are labeled as e, e_1 and e_2 respectively.

8.3.1 Simple structures

For the simple structures, results using different eigenvalues are the same for the simple block and simple intermediate structures. The simple block structure was retrieved exactly regardless of eigenvalues. But SCA2 applied to any eigenvalues did not get the simple intermediate structure back. This is expected from the results in Section 4.6.3 for the simple intermediate structure. So, eigenvalues don't influence the results for simple block and simple intermediate structures, and the same transformation matrix at the same step for the same structure leads to the same results regardless of the eigenvalues used.

For simple uniform structure, SCA2 (Figure 8.5) gets the same results when the eigenvalues e and e_2 are used, either way SCA2 gets the simple uniform structure back exactly, in order to distinguish them, I add 2° for each angle for SCA2 using $e2$. The results obtained by SCA2 using e_1 are the same for the first component but much less accurate than those obtained by SCA2 using e and e_2 for the second component. If SCA2 using e took steps 1, 2, 3, 4, 5, 6, 7, then using e_1 the steps were 1, 2, 3, 4, 5*, 6*, 7 (i.e. steps 5 and 6 are different when using e and e_1) and with e_2 the steps were 1, 2, 3, 4, 6, 7, 5. The steps taken when e and e_2 were specified led to the results being the same because SCA2 with e updates columns 1 and 4 in step 5, updates columns 5 and 6 in step 6 and updates columns 2 and 3 in step 7. All of these pairs are different, so the multiplication of the three transformation matrices in steps 5, 6 and 7 is the same regardless of the order. The transformation steps 5* and 6* updates different columns from transformation steps 5 and 6. So SCA2 using e_1 gets different results from SCA2 using e and e_2.

For the simple uniform structure, all the results (Figure 8.6) are simple, the results obtained by SCA2 using e and e_2 are more simple than those using e_1, in order to distinguish them, I add 0.1 for each mean of components obtained by SCA2 using $e2$. So the best results are obtained by SCA2 using e and e_2.

All in all, for simple block and simple intermediate structures, the results obtained by SCA2 are the same regardless of the eigenvalues. The results obtained by SCA2 with e_2 are more

accurate than those with e_1 for simple uniform structure. The number of steps taken by SCA2 is the same for the same structure of eigenvector for all the three sets of eigenvalues.

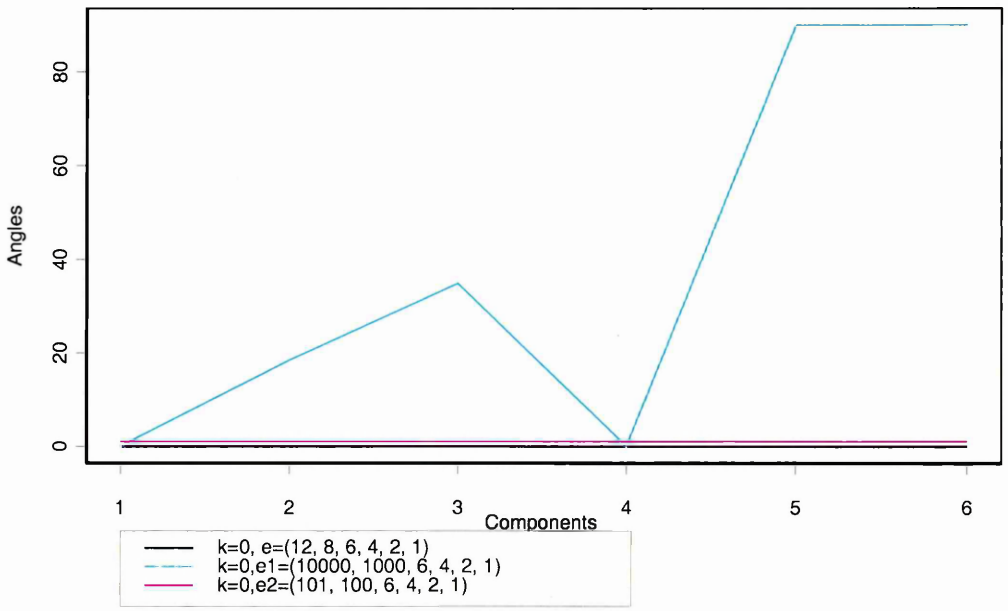

Figure 8.5: The accuracy of components produced by SCA2 using different eigenvalues for simple uniform structure

8.3.2 Complex structures

For the complex structures, the results, both in terms of accuracy and simplicity, are the same regardless of the eigenvalues chosen for the first two components for complex block and complex intermediate structures. This is surprising. The results of SCA2 using e and e_2 retrieved the corresponding simple block structure exactly. So the first two angles using different eigenvalues for the complex block structure are the angles between the first two components of simple block structure and the first two components of the complex block structure. But for the complex intermediate structure, all of the results are not as good as those for the complex block

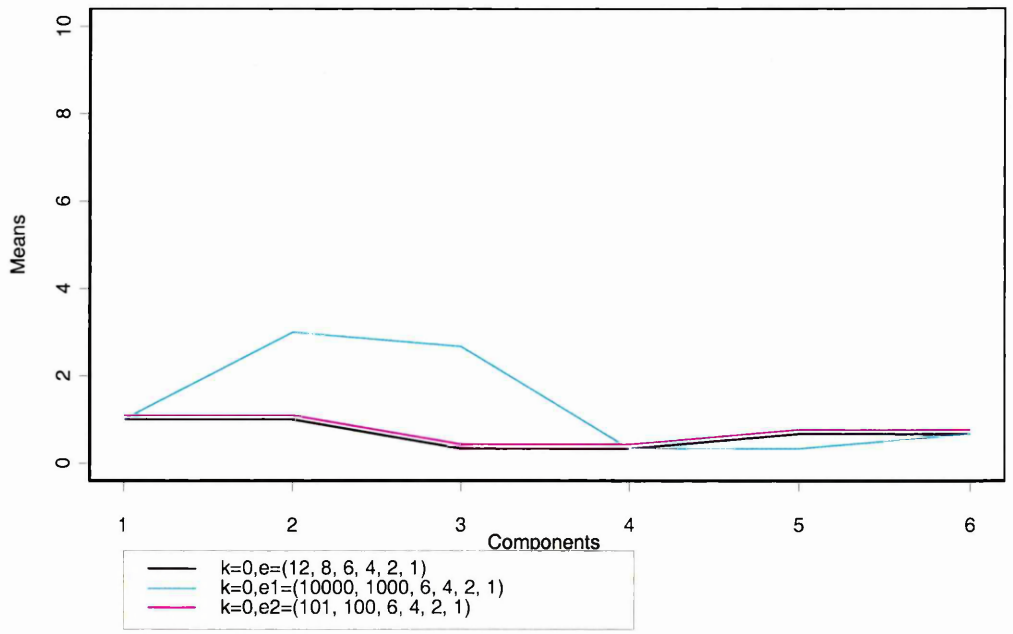

Figure 8.6: The simplicity of components produced by SCA2 using different eigenvalues for simple uniform structure

and complex uniform structures. This is expected from the results of the simple intermediate structure in Section 8.3.1 and the results in Section 4.6.3, and it is known from Section 8.2.2 that the most accurate results for complex intermediate structure are obtained when $k = 1$.

For the complex uniform structure, the results (Figure 8.7) using e_1 are the most accurate, the results using e are the second most accurate, the results using e_2 are the least accurate. This is much different from the results for simple uniform structure. SCA2 using e_1 retrieves the first two components of the corresponding simple uniform structure, so the first two angles for the complex uniform structure are just the angles between the first two components of complex uniform structure and simple uniform structure. SCA2 using e just retrieves the first component of the simple uniform structure.

For the complex uniform structure, the results (Figure 8.8) using e_2 are the simplest, the

results using e_1 are the second simplest, the results using original eigenvalues, e, are the least simple. So the best results of the complex uniform structure are obtained by SCA2 applied to e_1.

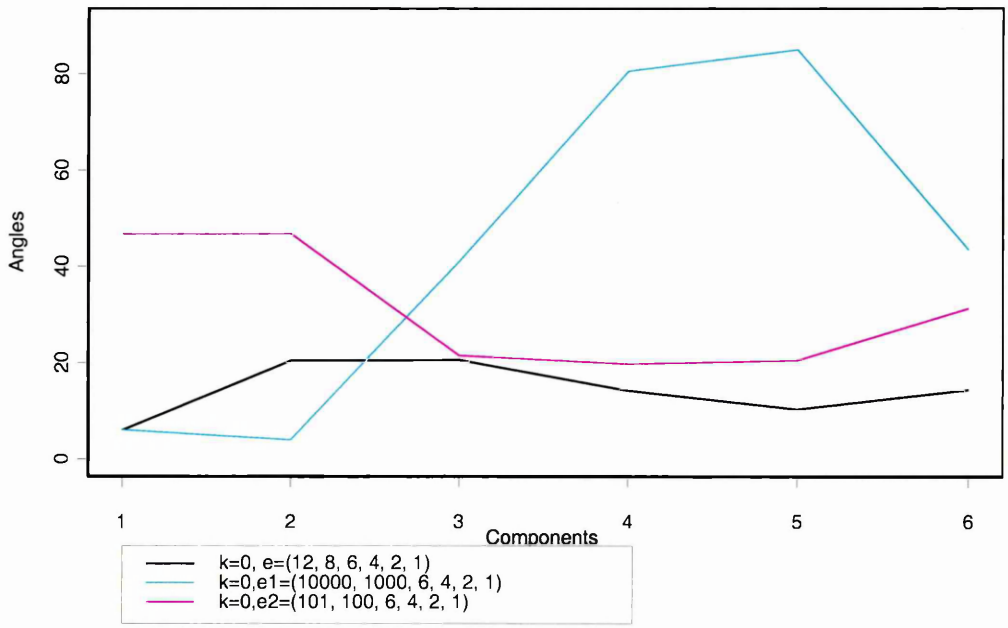

Figure 8.7: The accuracy of components produced by SCA2 using different eigenvalues for complex uniform structure

To sum up, for the complex block and complex intermediate structures, the results of SCA2 using all the eigenvectors are the same, for complex uniform structure, the results of SCA2 using e_1 are more accurate than those using e_2. So for the complex structures, in generally the results of SCA2 using e_1 are not less accurate than those using e_2. This is expected considering the problems of similar eigenvalues in PCA (Jolliffe (2002b) Sections 2.4, 3.4, 3.7 and 10.3). The simple components are very close to the corresponding principal components, so the same problems of principal components also apply. This is different from simple structures. Again, the eigenvalues do not affect the number of steps SCA2 taken for the same structure.

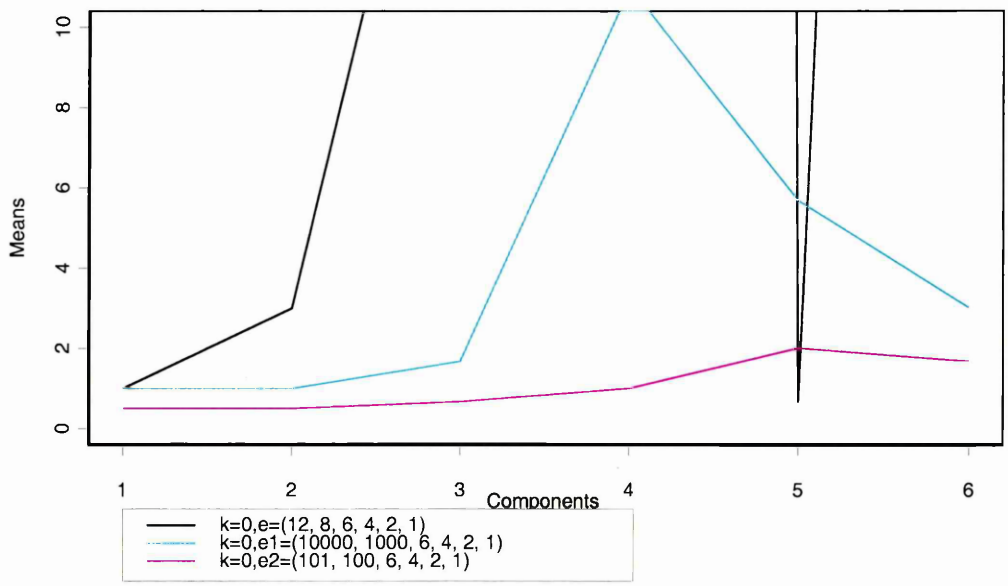

Figure 8.8: The simplicity of components produced by SCA2 using different eigenvalues for complex uniform structure

8.4 Results for different k and different eigenvalues for large dimensional data

In this section, the results for different k and different eigenvalues will be investigated for 8 and 10 dimensional data. First, in Section 8.4.1, I discuss the effect of k on population results obtained by SCA2 as k increases.

8.4.1 Effect of k

The results obtained by SCA2 with larger k for large dimensional data generally are no more accurate and not as simple as those applied to smaller k. This is the same as what has been observed in Section 8.2 for 6 dimensional data. So only the results for the simple intermediate

structures are discussed.

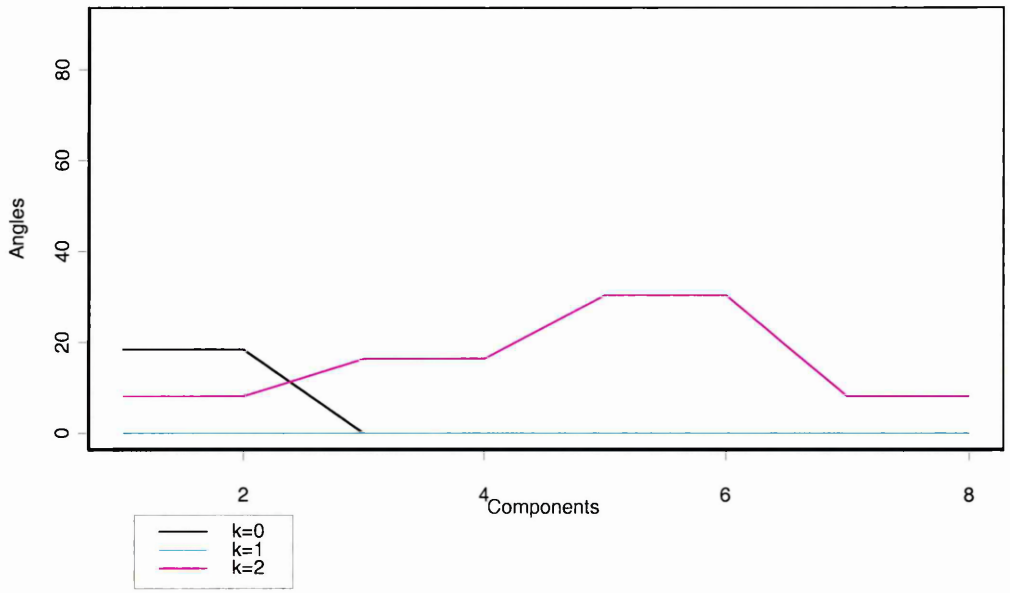

Figure 8.9: The accuracy of components produced by SCA2 using different k for 8 dimensional simple intermediate structure

For the 8 dimensional simple intermediate structure, SCA2 retrieves the 8 dimensional simple intermediate structure exactly when $k = 1$ (Figure 8.9). These are the most accurate results. The next most accurate results for the first two components are obtained when $k = 2$. The least accurate results are obtained when $k = 0$. As k increases, the results are not always more accurate.

The simplest results (Figure 8.10) are obtained by SCA2 when $k = 0$, so it is too simple considering that when $k = 1$ SCA2 retrieves the simple intermediate structure back exactly. Next most simple results are obtained when $k = 1$. Generally the smaller k, the more simple the results produced by SCA2. So for 8 dimensional simple intermediate structure, SCA2 using $k = 1$ gets the best results.

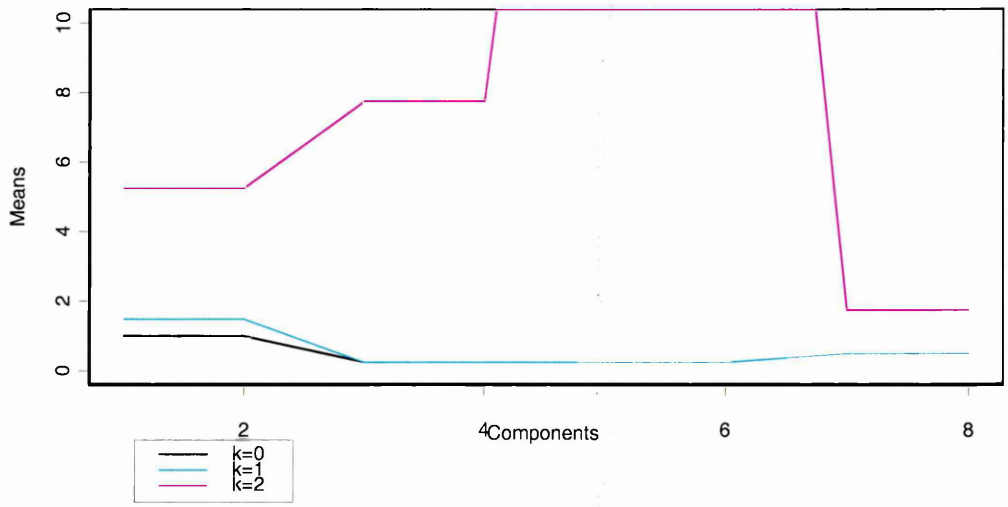

Figure 8.10: The simplicity of components produced by SCA2 using different k for 8 dimensional simple intermediate structure

For the 10 dimensional simple intermediate structure, the results obtained by SCA2 when $k = 1$ are not very good. However when $k = 1$ SCA5 and SCA6 retrieve the 10 dimensional simple structure.

Why can the SCA methods retrieve the simple intermediate structure when $k = 1$? This is because the loadings of the transformation matrix at each step are more likely to be similar to the loadings of the simple intermediate structures in Sections 4.3 and 6.2 when $k = 1$.

In general, the results obtained by SCA2 applied to larger k are not always more accurate and not as simple as those applied to smaller k whatever the dimension of the data.

8.4.2 The results using different sets of eigenvalues

In this subsection, I am going to investigate the results of SCA2 using different sets of eigenvalues for large dimensional data. For 8 dimensional data, two new sets of eigenvalues (30000, 200, 14, 12, 10, 6, 4, 2) and (2001, 2000, 14, 12, 10, 6, 4, 2) are investigated. For 10 dimensional data, two new sets of eigenvalues (15000, 500, 16, 14, 12, 10, 8, 6, 4, 2) and (101, 100, 16, 14, 12, 10, 8, 6, 4, 2) are investigated.

For the larger dimensional data, whether having a simple or complex structure, SCA2 took the same number of steps for the same structure regardless of the eigenvalues. This is the same as what obtained for 6 dimensional data. The results obtained by SCA2 using eigenvalues with bigger difference generally are better than those with little difference for large dimensional data. For 6 dimensional data, the results with bigger difference generally are the same or more accurate than those with smaller difference. So in general the results obtained by SCA2 using eigenvalues with bigger difference generally are better than those with little difference whatever the dimension of the data. This is expected considering the problems of similar eigenvalues in PCA (Jolliffe (2002b) Sections 2.4, 3.4, 3.7 and 10.3). The simple components are very close to the corresponding principal components, so the same problems of principal components also apply.

8.5 Discussion and conclusion

In Chapter 8, the population results obtained by SCA2 using different k ($k = 0$, 1 and 2) and using three different sets of eigenvalues were investigated. It is shown that the results obtained by SCA2 with larger k are not necessarily more accurate than those with smaller k. This is because generally the maximal improvement in variance at a given step just makes the results the most accurate at that step. Furthermore, the maximal improvement in variance of SCA2 with larger k at each step is not always greater than the maximal improvement in variance with smaller k because after some step the variance-covariance matrix is different for different k (an example was given in Section 8.2.1). Even if the maximal improvement in variance obtained by SCA2 with large k is more than that obtained by SCA2 with smaller k at each step, this can

not guarantee the most accurate results finally (an example of this was given in Section 8.2.1). All the above conclusions are also applied to other SCA methods.

Furthermore the results with larger k are not as simple as those with smaller k. This is because the larger the k, the longer the new directions in general (Sections 3.3.2, 3.3.3 and 8.2.1). Also as k increases, generally the steps taken by SCA2 increase and more steps also mean that the components are longer generally. So the results of SCA methods with larger k are usually not better than those with smaller k because they are not as simple as those with smaller k.

It was shown in Chapters 4 to 7 that when $k = 0$, SCA was not very good for intermediate structure. However in this chapter, it was shown that when $k = 1$ SCA2 generally retrieved the first two components of simple intermediate structure or the simple intermediate structure. The reason is that when $k = 1$, the loadings of the transformation matrix of SCA2 for 2 dimensional data at each iteration are more likely to be similar to the loadings of the simple intermediate structure. For more than two dimensional data, the loadings of the multiplication of the transformation matrices are more likely to be similar to the loadings of simple intermediate structure. So SCA methods can get the first two components of all the simple structures back exactly if a suitable k is chosen. The angles obtained by SCA2 for the complex intermediate structure were just the angles between the first two components of simple intermediate structure and complex intermediate structure when $k = 1$. These conclusions are the same as those for block and uniform structures when $k = 0$. For complex structures the best angles generally are obtained by SCA methods when the SCA method retrieves the simple structure because by definition the simple structure is the closest simple approximation of the corresponding complex structure.

In Chapters 5 and 7, for simple block and simple uniform structures when $k = 0$ it was concluded that the sample simulation results obtained by SCA are better than those obtained by PCA. So I conclude that the sample simulation results obtained by SCA methods should be better than the sample simulation results obtained by PCA for simple intermediate structure when $k = 1$ (the reason is similar to what was given in Section 5.5). I have done the sample simulation when $k = 1$ for 6 dimensional intermediate structure in Section 4.3, this is true

for 6 dimensional intermediate structure. So for any simple structure, the sample simulation results obtained by the SCA algorithm should be better than those obtained by PCA if a suitable k is used. That is to say, whatever the dimension for simple structures the sample simulation results obtained by SCA methods should not only be easier to interpret but also more accurate than those obtained by PCA. If an individual SCA method gets the simple intermediate structure for both simple and complex intermediate structures, and the angles between the simple intermediate structure and complex intermediate structure are smaller than the difference between the sample simulation results between PCA and this SCA method for simple intermediate structure when $k = 1$, the sample simulation results of the corresponding complex intermediate structures are better than those of PCA when $k = 1$. This means that whatever the structure of the data, if PCA is good for some data, SCA is also generally good so long as a suitable k is chosen. In general, the best results appear to be obtained when k equals 0 or 1.

It is also shown in Chapter 8 that generally the eigenvalues of a data set influence the results, but not by very much if the same eigenvectors are used. For the simple block structure all the results for the first two components were the same. So it seems that the eigenvalues of the variance-covariance matrix had no effect on the results of simple block structure, though surely this is not true for all possible choices of eigenvalues. For the other structures, the results of SCA methods applied to the variance-covariance matrices that had different eigenvalues generally differed but not by very much. In general, when there is a big difference between the first two eigenvalues, the results were more accurate than those obtained when there was an extremely small difference between the first two eigenvalues. This is expected. If the first two eigenvalues have an extremely small difference, the first two principal components have almost equal variance. It is known that the subspace spanned by the eigenvectors which have exactly equal eigenvalues that are well-separated from all other eigenvectors is well defined and stable. But the eigenvectors themselves are unstable and are not uniquely defined (Jolliffe (2002b) Sections 2.4, 3.4, 3.7 and 10.3). So if the first two eigenvalues have an extremely small difference because of rounding error, the population results are sometimes unstable. When there is a big difference between the first two eigenvalues, all the eigenvalues are different because all the other

eigenvalues are assumed to be the same as the eigenvalues used in previous chapters. So PCA generally retrieves the corresponding eigenvectors for population variance covariance matrices, and for sample simulation data, PCA is very good. The simple components are very close to the corresponding principal components, so the above properties of principal components also apply. This means that two simple components are unstable and not uniquely defined if the first two eigenvalues are exactly equal. The results obtained by SCA should be more accurate if the first two eigenvalues have a big difference. However the eigenvector structures are the main factor to affect the results of SCA methods.

Chapter 8 only investigated the effect of k and the effect of the sets of eigenvalues on the results obtained by SCA2. Of course the same issues can be investigated for all the other SCA methods, but considering the small difference between the different SCA methods, it is expected that all of the conclusions obtained by SCA2 can be extended to other SCA methods. Also in Chapter 8, only population results were investigated. Of course, sample simulation results can also be considered. It is expected that sample simulation should get similar conclusions to the population results.

It is known that eigenvectors associated with the data are a major factor to affect the results of SCA methods. The same type of structures but with different eigenvectors from the data sets in Sections 4.3 and 6.2 with same or different eigenvalues can also be investigated. The general conclusions about SCA methods should be similar but maybe some new features such as the effect of eigenvectors on the SCA methods would be found.

Chapter 9

The population results obtained by SCA2 using different restrictions

9.1 Introduction

As said in Section 3.3, in order to get simple components, the directions considered for each simple preserving transformation are limited by the restrictions (3.10) and (3.11). May other restrictions on the directions available be used to get simple components? Are the results obtained by SCA methods using new restrictions improved compared to the results obtained by SCA methods using restrictions (3.10) and (3.11)? To try to answer these questions, in this chapter two new restrictions to the directions available will be introduced. As Chapters 4 to 7 showed, SCA2 is one of the best methods among all the SCA methods. Also whatever the dimension of the data, the conclusions about the SCA methods are the same whether the population results are investigated or the sample simulation results are investigated. So, only the population results obtained by SCA2 using three different restrictions will be considered in this chapter, although I have investigated the results of all the other SCA methods using different restrictions.

In Section 9.2, two new restrictions to the directions available for SCA methods will be introduced. In Section 9.3, the directions available using three different restrictions when $k = 0$,

1 and 2 are given by graphs. In Section 9.4, I will compare the population results obtained by SCA2 using different restrictions for 6 dimensional data when $k = 1$. In Section 9.5, I will investigate the population results obtained by SCA2 using different restrictions for 8 and 10 dimensional data when $k = 1$. The last section, Section 9.6, is the discussion and conclusion.

9.2 Different restrictions

As in Section 3.3.1, suppose d_1 and d_2 represent the two updated directions in iteration 1. Also assume that the lengths of d_1 and d_2 are 1 at iteration 0. It is known from Section 3.3 that in general the linear transformation of the SCA algorithm can be represented as follows. In equation (3.9),

$$\left.\begin{aligned} f_1 &= d_1 + \beta d_2 \\ f_2 &= l_2^2 \beta d_1 - l_1^2 d_2 \end{aligned}\right\}$$

for some suitable choice of β.

Generally, f_1 and f_2 are not integer vectors even if d_1 and d_2 are. As said in Section 3.3, this problem can be overcome by restricting β to be either $b = \frac{i}{2^k}$ when $|b| \leq 1$ or $b = \frac{2^k}{i}$ when $|b| > 1$, $i = -2^k, -2^k + 1, \ldots, 2^k$ (Vines, 2000), where k is a nonnegative integer. So we get the restriction (3.10) if we substitute in values of b for which $|b| \leq 1$, and get restriction (3.11) if we substitute in values of b for which $|b| > 1$. So restriction (3.10) is

$$\left.\begin{aligned} f_1 &= 2^k d_1 + i d_2 \\ f_2 &= i l_2^2 d_1 - 2^k l_1^2 d_2 \end{aligned}\right\} |b| \leq 1$$

and restriction (3.11) is

$$\left.\begin{aligned} f_1 &= i d_1 + 2^k d_2 \\ f_2 &= 2^k l_2^2 d_1 - i l_1^2 d_2 \end{aligned}\right\} |b| > 1$$

The performance of the SCA algorithm explored in Chapters 4 to 8 is based on restrictions (3.10) and (3.11). For convenience, the restrictions (3.10) and (3.11) are represented as restriction I. Restriction I just guarantee that f_1 and f_2 are integers if d_1 and d_2 are, it does not guarantee that the best results are obtained by SCA methods. May we get other restrictions to ensure that f_1 and f_2 are integer vectors if d_1 and d_2 are, and also ensure that the results of SCA methods using the new restrictions are improved? So the following two new restrictions are introduced.

The first new restriction is obtained by restricting β in equation (3.9) to be either $2^k b$ when $|b| \leq 1$, or $\frac{b}{2^k}$ when $|b| > 1$, where k is a nonnegative integer and $b = \frac{i}{2^k}$ when $|b| \leq 1$ or $b = \frac{2^k}{i}$ when $|b| > 1$, $i = -2^k, -2^k + 1, \ldots, 2^k$. i.e. when $|b| \leq 1$, the β in equation (3.9) is equal to i, when $|b| > 1$, the β in equation (3.9) is equal to $1/i$.

$$\left.\begin{aligned} \mathrm{d}_1^* = \mathrm{f}_1 &= d_1 + id_2 \\ \mathrm{d}_2^* = \mathrm{f}_2 &= il_2^2 d_1 - l_1^2 d_2 \end{aligned}\right\} \; |b| \leq 1 \tag{9.1}$$

and

$$\left.\begin{aligned} \mathrm{d}_1^* = \mathrm{f}_1 &= id_1 + d_2 \\ \mathrm{d}_2^* = \mathrm{f}_2 &= l_2^2 d_1 - il_1^2 d_2 \end{aligned}\right\} \; |b| > 1 \tag{9.2}$$

This ensures that d_1^* and d_2^* will be integer vectors whenever d_1 and d_2 are integer vectors. Restrictions (9.1) and (9.2) are called as restriction II.

A second new restriction can be formed by restricting β in equation (3.9) to be either $b = \frac{i}{3^k}$ when $|b| \leq 1$ or $b = \frac{3^k}{i}$ when $|b| > 1$, $i = -3^k, -3^k + 1, \ldots, 3^k$, where k is non-negative integers. This means, setting

$$\left.\begin{aligned} \mathrm{d}_1^* = \mathrm{f}_1 &= 3^k d_1 + id_2 \\ \mathrm{d}_2^* = \mathrm{f}_2 &= il_2^2 d_1 - 3^k l_1^2 d_2 \end{aligned}\right\} \; |b| \leq 1 \tag{9.3}$$

$$\left.\begin{aligned} \mathrm{d}_1^* = \mathrm{f}_1 &= id_1 + 3^k d_2 \\ \mathrm{d}_2^* = \mathrm{f}_2 &= 3^k l_2^2 d_1 - il_1^2 d_2 \end{aligned}\right\} \; |b| > 1 \tag{9.4}$$

This ensures that d_1^* and d_2^* will be integer vectors whenever d_1 and d_2 are integer vectors. The value $b = \beta$ which maximizes v is then sought within this restricted set. Restrictions (9.3) and (9.4) are represented as restriction III.

For all the restrictions, generally, d_1^* will tend to be a more simple vector than d_2^*. This is because all the restrictions can be written as

$$\left.\begin{aligned} \mathrm{d}_1^* &= \mathrm{f}_1 = cd_1 + id_2 \\ \mathrm{d}_2^* &= \mathrm{f}_2 = il_2^2 d_1 - cl_1^2 d_2 \end{aligned}\right\} |b| \leq 1 \tag{9.5}$$

$$\left.\begin{aligned} \mathrm{d}_1^* &= \mathrm{f}_1 = id_1 + cd_2 \\ \mathrm{d}_2^* &= \mathrm{f}_2 = cl_2^2 d_1 - il_1^2 d_2 \end{aligned}\right\} |b| > 1 \tag{9.6}$$

where c is a constant for fixed k. In restriction I, $c = 2^k$, in restriction II, $c = 1$ and in restriction III, $c = 3^k$. So when $k = 0$, all the three restrictions are the same. So Under restriction (9.5),

$$d_1^{*T} d_1^* = (cd_1 + id_2)^T (cd_1 + id_2) = (c^2 l_1^2 + i^2 l_2^2)$$

$$d_2^{*T} d_2^* = (il_2^2 d_1 - cl_1^2 d_2)^T (il_2^2 d_1 - cl_1^2 d_2) = l_1^2 l_2^2 (c^2 l_1^2 + i^2 l_2^2).$$

And under restriction (9.6),

$$d_1^{*T} d_1^* = (id_1 + cd_2)^T (id_1 + cd_2) = (i^2 l_1^2 + c^2 l_2^2)$$

and

$$d_2^{*T} d_2^* = (cl_2^2 d_1 - il_1^2 d_2)^T (cl_2^2 d_1 - il_1^2 d_2) = l_1^2 l_2^2 (i^2 l_1^2 + c^2 l_2^2)$$

where d_1 and d_2 are integer vectors. Now as said in Section 3.3.1, $l_1^2 l_2^2 \geq 1$ at any iteration. So, in general, the length of d_2^* is greater than that of the d_1^* under all the restrictions, i.e d_1^* is more simple than d_2^*.

Dividing the corresponding lengths using restrictions I and III, we get $rl = (\frac{2}{3})^k \sqrt{\frac{l_1^2+b_1 l_2^2}{l_1'^2+b_2 l_2'^2}}$, where l_1 and l_2 represent the lengths of the updated two directions using restriction I and l_1' and l_2' represent the lengths of the updated two directions using restriction III, b_1 and b_2 represent the values of b using restrictions I and III respectively. When $k = 0$, the lengths using restrictions I and III are the same because they are the same restrictions. When $k \geq 1$ the components produced by SCA methods using restriction I generally are more simple than those produced by SCA methods using restriction III at each iteration. I just explain this at iteration 1, the lengths of updated two directions in other iterations can be interpreted in a similar way. Recall that at iteration 0, $l_1 = l_1' = l_2 = l_2' = 1$, the possible value of b using restriction I is 0, $\pm 1/2$, ± 1, ± 2, the possible value of b using restriction III is 0, ± 1, $\pm 1/3$, $\pm 2/3$, $\pm 3/2$ and ± 3. So when $k = 1$ and $b_1 = b_2$, $r1 = 2/3 < 1$. When $k = 1$ and $b_1 \neq b_2$, the possible values of b_1 are $\pm 1/2$ and ± 2, the possible values of b_2 are $\pm 1/3$, $\pm 2/3$, $\pm 3/2$, and ± 3, in this case, the maximal value of $r1 = \frac{2}{3}\sqrt{\frac{1+2}{1-2/3}} = 2$. The second maximal value of $r1 = \frac{2}{3}\sqrt{\frac{1+2}{1-1/3}} = \frac{2}{3}\sqrt{\frac{1+1/2}{1-2/3}} = \sqrt{2}$. The third maximal value of $r1 = \frac{2}{3}\sqrt{\frac{1+2}{1+1/3}} = 1$. In other cases, $rl < 1$. So when $k = 1$, at iteration 1 the components produced by SCA methods using restriction I generally are more simple than those produced by SCA methods using restriction III.

All in all, for all the three restrictions, d_1^* is more simple than d_2^*. If $k = 0$, all the restrictions are the same. If $k \geq 1$, in general, the lengths of the updated two directions using restriction III are not as simple as those using restriction I at each iteration.

9.3 Allowable directions for the pairwise transformations using different restrictions

In this section, the allowable directions for SCA methods using different restrictions will be investigated.

First, graphs of allowable directions using different restrictions and different value of k when $l_1 = l_2$ are considered.

Figure 9.1 shows that when $k = 0$, the directions available are the same for all the three

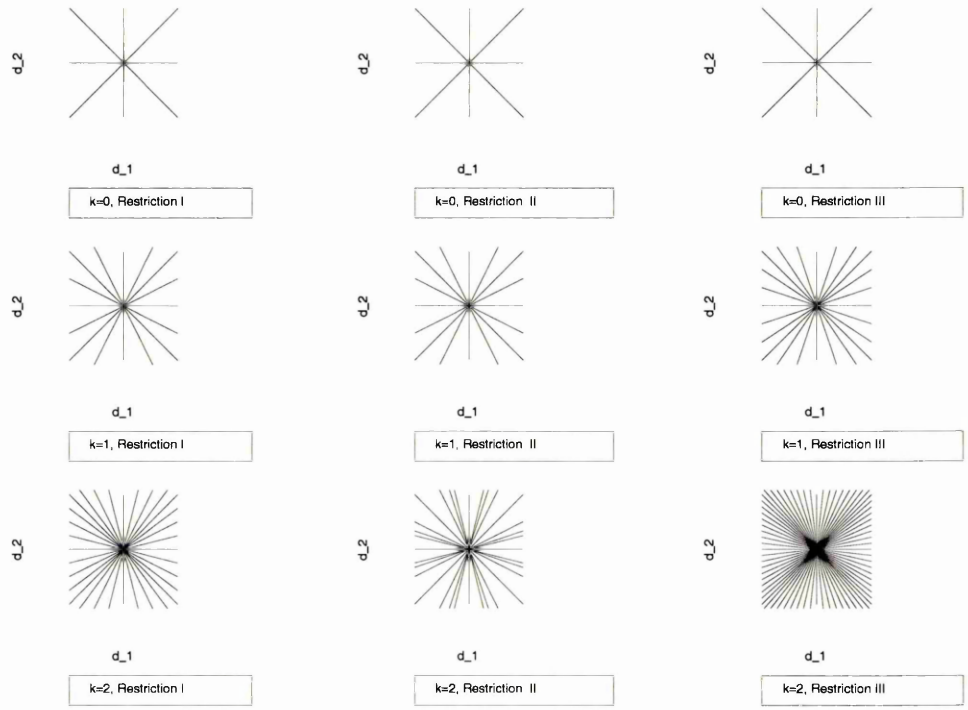

Figure 9.1: The allowable directions when $l_1 = l_2$

restrictions. This is because as said in Section 9.2, when $k = 0$ for all the three restrictions c is 1 in restrictions (9.5) and (9.6). When $k = 1$ and $k = 2$, the number of the directions available using restriction III is more than those available using restrictions I and II. For example, when $k = 1$, d_1^* using restriction III corresponds to one of 12 possible directions $d_1 - 3d_2$, $d_1 - \frac{3}{2}d_2$, $d_1 - d_2$, $d_1 - \frac{2}{3}d_2$, $d_1 - \frac{d_2}{3}$, d_1, $d_1 + \frac{d_2}{3}$, $d_1 + \frac{2}{3}d_2$, $d_1 + d_2$, $d_1 + \frac{3}{2}d_2$, $d_1 + 3d_2$ and d_2, but d_1^* using restrictions I and II corresponds to one of only 8 possible directions (the 8 directions were given in Section 3.3.2). However, when $k = 1$, for restrictions I and II the same value of b corresponds to different directions. For example, when $b = 1$, the directions corresponding to restriction I are $d_1 + d_2$ and $d_1 - d_2$, the directions corresponding to restriction II are $d_1 + 2d_2$ and $2d_1 - d_2$. This is surprising, I suppose restrictions I and II are different restrictions when $k = 1$, they look so different when $k = 1$. That is to say restrictions I and II are identical when $k = 0$ and 1. When $k = 2$, the numbers of the directions using these two restrictions are the same but directions

available generally are different. As k increases, the number of the directions available using the same restriction increases, and the difference between the number of the directions using restrictions I and II and number of directions using restriction III increases. When $k = 1$, the directions available using restriction I and restriction II are the same. When $k = 2$, the number of the directions using restriction I is the same as that using restriction II but the directions available generally are different from those using restriction II. When $k = 2$, d_1^* using restriction III corresponds to one of 36 possible directions, but d_1^* using restrictions I and II corresponds to one of only 16 possible directions.

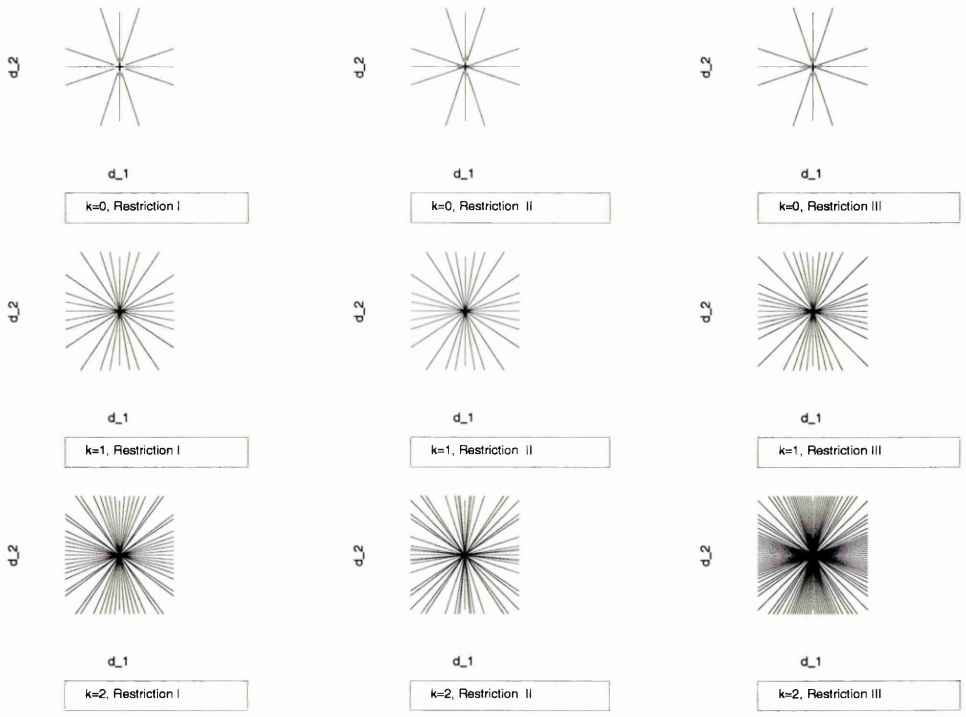

Figure 9.2: The allowable directions when $\frac{l_1}{l_2} = 3$

Next consider the directions using different restrictions when $l_1/l_2 = 3$ and $k = 0$, 1 and 2. Figure 9.2 shows that again, when $k = 0$ (row 1), the directions available are the same for all the restrictions, and again as k increases, the number of the directions using the same restriction increases. When $k = 1$ and 2, the number of allowable directions using restriction III is the

most and the number of directions using the restriction I is the same as the number of directions using the restriction II. When $k = 0$ and 1, the directions using restrictions I and II are identical. When $k = 2$, the number of directions using restriction II is the same as that using restriction I, but the directions available generally are different.

Finally consider the allowable directions at each iteration using different restrictions (Figure 9.3), when $\frac{l_1}{l_2} = 5$ and $k = 0$, 1 and 2. The conclusions when $\frac{l_1}{l_2} = 5$ are the same as those when $\frac{l_1}{l_2} = 3$, but the corresponding directions for a given restriction when $\frac{l_1}{l_2} = 5$ are more steep.

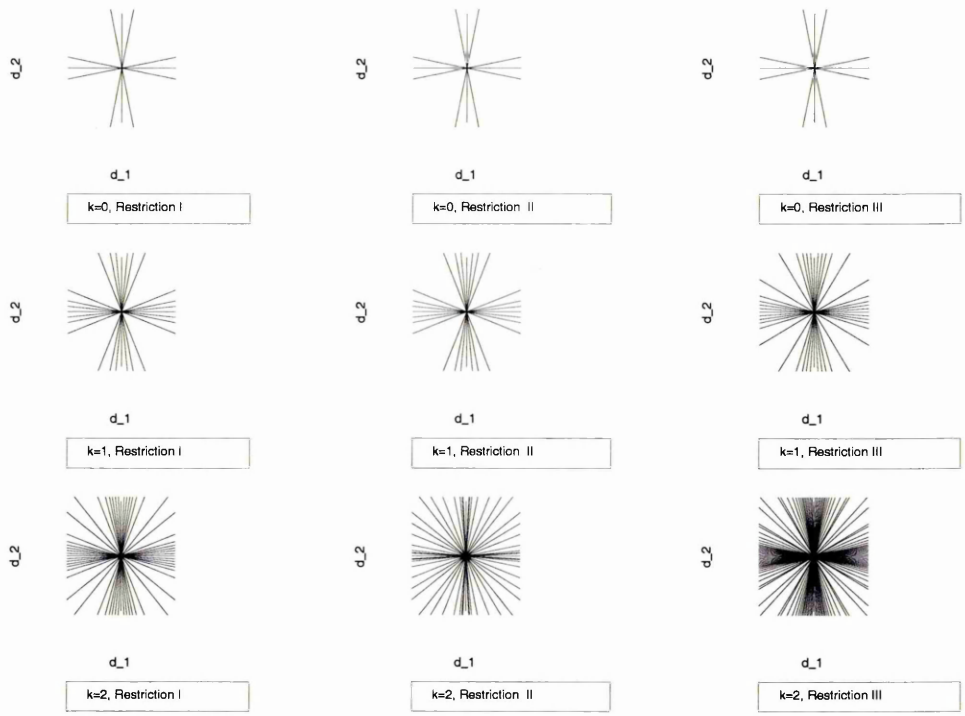

Figure 9.3: The allowable directions when $\frac{l_1}{l_2} = 5$

Comparing the allowable directions using different restrictions when $\frac{l_1}{l_2} = 1$, 3 and 5 (Figures 9.1, 9.2 and 9.3) when $k = 0$, 1 and 2, the numbers of the directions when $\frac{l_1}{l_2} = 3$ and 5 are the same in all the cases for a given k. When $\frac{l_1}{l_2} = 1$, for a given restriction, there are fewer directions available (Figure 3.1) because as mentioned before in Section 3.3.3, in this case some directions become the same. As $\frac{l_1}{l_2}$ increases, for a given restriction, some of the directions are nearer the

x axis and some of the directions nearer y axis compared to the corresponding directions when $\frac{l_1}{l_2} = 1$.

So when $k = 0$, all the restrictions are the same. When $k = 1$ restrictions I and II are identical. Restrictions I and II are different when $k = 2$. When $k \geq 1$, the number of directions available using restriction III is more than those using restrictions I and II, so the results obtained by SCA methods using restriction III should be more accurate than those using restrictions I and II. But, in Chapter 8, we concluded that the results obtained by SCA methods with larger k (i.e. more directions) are not always more accurate than those obtained by SCA methods with smaller k, so the results obtained by SCA methods using restriction III might be more accurate, or might not be more accurate than those obtained by SCA methods using restrictions I and II. Furthermore the two updated directions using restriction III are not as simple as the two updated directions using restrictions I and II when $k = 1$ at each iteration (the reason was given in Section 9.2). So the results obtained by SCA methods using restriction III might not be as simple as those using restrictions I and II. That is to say more accurate and less simple results generally are expected for SCA methods using restriction III compared to the results obtained by SCA methods using restrictions I and II.

9.4 Population results for 6 dimensional data-$k = 1$

From Sections 9.2 and 9.3, it is known that when $k = 0$ the restrictions I, II and III are identical, so the results obtained by SCA methods using restrictions I, II and III are the same. So, in order to compare the different restrictions, SCA2 using different restriction when k is greater than zero will be investigated. However, it is also known that generally SCA methods are more likely to get the best results when $k = 0$ or 1 (Section 8.5), and restrictions I and II are also identical when k is equal to 1. So in this section, I just compare the results obtained by SCA2 using restrictions I and III when $k = 1$. SCA2 using restrictions I and III when $k = 1$ are called SCA2I and SCA2III respectively.

9.4.1 Simple structures

Consider the population results obtained by SCA2 using restrictions I and III for the simple structures defined in Section 4.3.

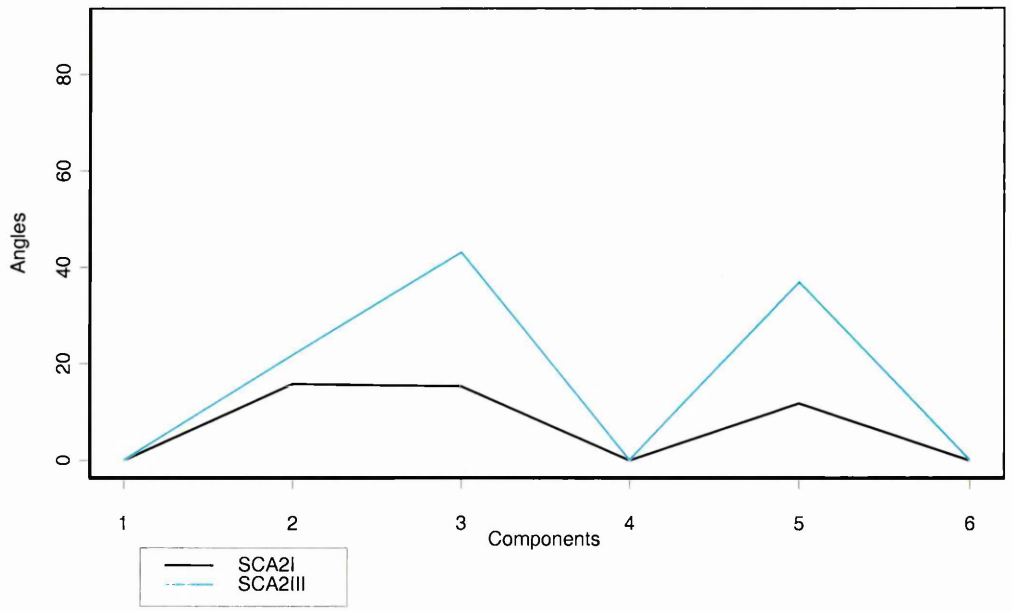

Figure 9.4: The accuracy of components produced by SCA2I and SCA2III for 6 dimensional simple block structure

For the simple uniform structure, the results obtained by SCA2III are much more accurate than those obtained by SCA2I. This is just what was expected. But this is not true for the other two simple structures. For the simple block structure, SCA2 using restrictions I and III both get the first component back exactly (Figure 9.4). For the second component, the component produced by SCA2I is more accurate than that produced by SCA2III, moreover as Figure 9.4 shows, all the components produced by SCA2I were at least as accurate as those produced by SCA2III. So the results obtained by SCA2 using restriction I are more accurate than those obtained by SCA2 using restriction III. For the simple intermediate structure, SCA2

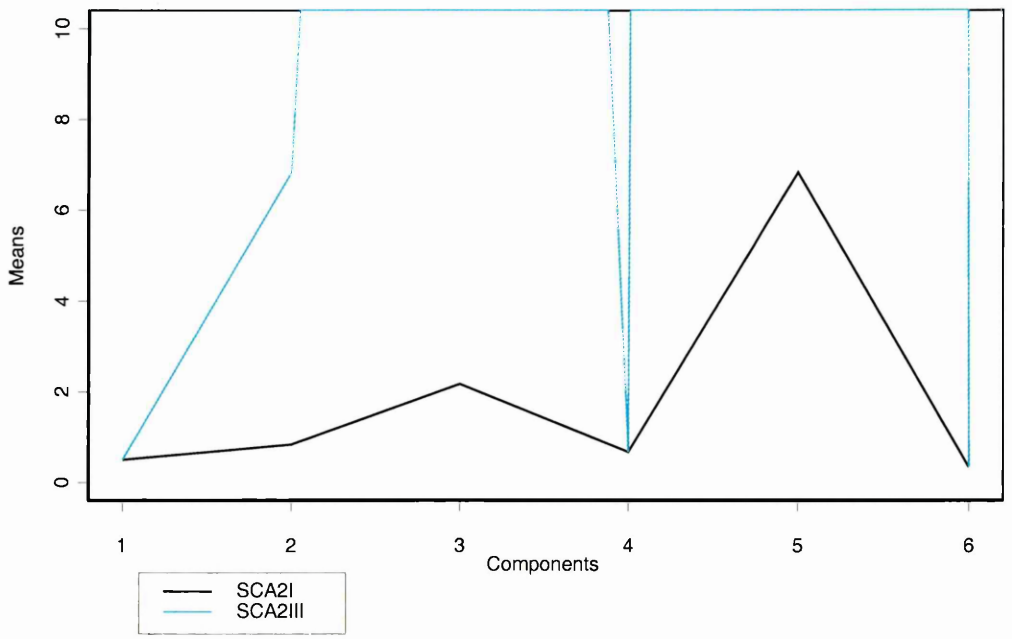

Figure 9.5: The simplicity of components produced by SCA2I and SCA2III for simple block structure

using restriction I gets the first two components back exactly. The results obtained by SCA2III are less accurate than those obtained by SCA2I for simple intermediate structure. So, in general, for simple structures, the results obtained by SCA2III are not always more accurate than those obtained by SCA2I. This is not what was expected. In Section 9.3 I suppose that the results obtained by SCA2III should be generally more accurate than those obtained by SCA2I because of the greater number of directions available for the simple preserving transformation matrix at each step for SCA2III. However it is not very surprising considering the results obtained by SCA2 as k increases (Section 8.2.1). The greater number of directions available for restriction III does not guarantee that at each step the maximal improvement in variance using restriction III is more than the maximal improvement in variance using restriction I. Sometimes the best maximal improvement in variance is obtained in a direction available using restriction I but not

using restriction III. Even if the maximal improvement in variance of SCA2 using restriction III at each step is more than that using restriction I, this does not guarantee more accurate results overall (an example of this was given in Section 8.2.1) because this does not guarantee the maximal improvement in variance overall.

For the first two components of the simple block structure, the results from restrictions I and III (Figure 9.5) are simple. However the results obtained by SCA2III are not as simple as those obtained by SCA2I for simple block structure. In fact, the results obtained by SCA2III are not as simple as those obtained by SCA2I for all the simple structures. This is not surprising. Just comparing the elements of the transformation matrix of SCA2III at each step with those of SCA2I at each step, in general the lengths of the allowable directions using SCA2III are longer than the corresponding lengths of directions using SCA2I at each step (the reason was given Section 9.2), and SCA2III generally took more steps than those taken by SCA2I. Recall that in Section 4.5.1 more steps taken by an individual SCA method generally mean less simple results produced by this method. So, in general, the components produced by SCA2III are less simple than those produced by SCA2I.

So the best results for 6 dimensional simple block and intermediate structures are obtained by SCA2I. The best results for simple uniform structure are obtained by SCA2III.

9.4.2 Complex structures for 6 dimensional data

In this subsection, we consider the accuracy of components produced by SCA2I and SCA2III for the complex structures for 6 dimensional data. For the complex block structure, the results obtained by SCA2I are more accurate than the results obtained by SCA2III (Figure 9.6). This is expected considering the results for simple block structure for 6 dimensional data. For the complex uniform structure, the results obtained by SCA2I are also more accurate than those of SCA2III. This is very surprising because for the simple uniform structure SCA2III got much more accurate results than SCA2I. This is because SCA2III did not retrieve the simple uniform structure both for the simple and complex uniform structures. So it is possible for SCA2I to get more accurate results than those of SCA2III for the complex uniform structure. For the simple intermediate structure, the results using the restrictions I and III are the same for the

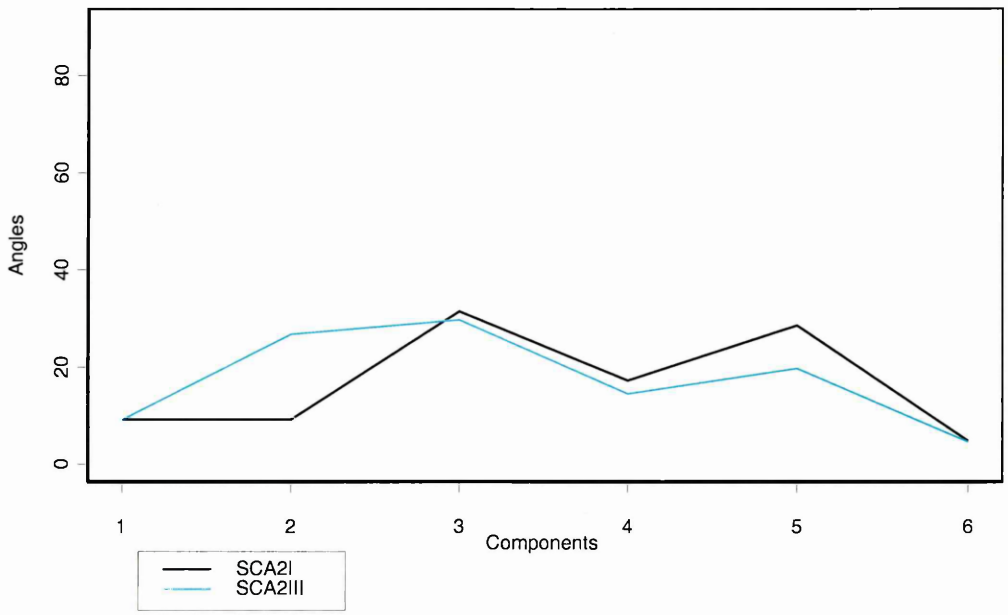

Figure 9.6: The accuracy of components produced by SCA2I and SCA2III for complex block structure

first two components. Furthermore, the first two angles are just the angles between the first two components of simple intermediate structure and the first two components of the complex intermediate structure and all the results are simple for the first two components. This is because SCA2I and SCA2III retrieve the first two components of the simple intermediate structure for the complex intermediate structure.

Next consider simplicity. The first two components produced by SCA2I and SCA2III (Figure 9.7) are simple for the complex block structure. The results obtained by SCA2I are more simple than those obtained by SCA2III. This is the same finding as was found for the simple block structure. In general, the results obtained by SCA2III are not as simple as those obtained by SCA2I for all the complex structures. These findings are the same as was found for simple block and simple intermediate structures and just what would be expected considering the analysis in

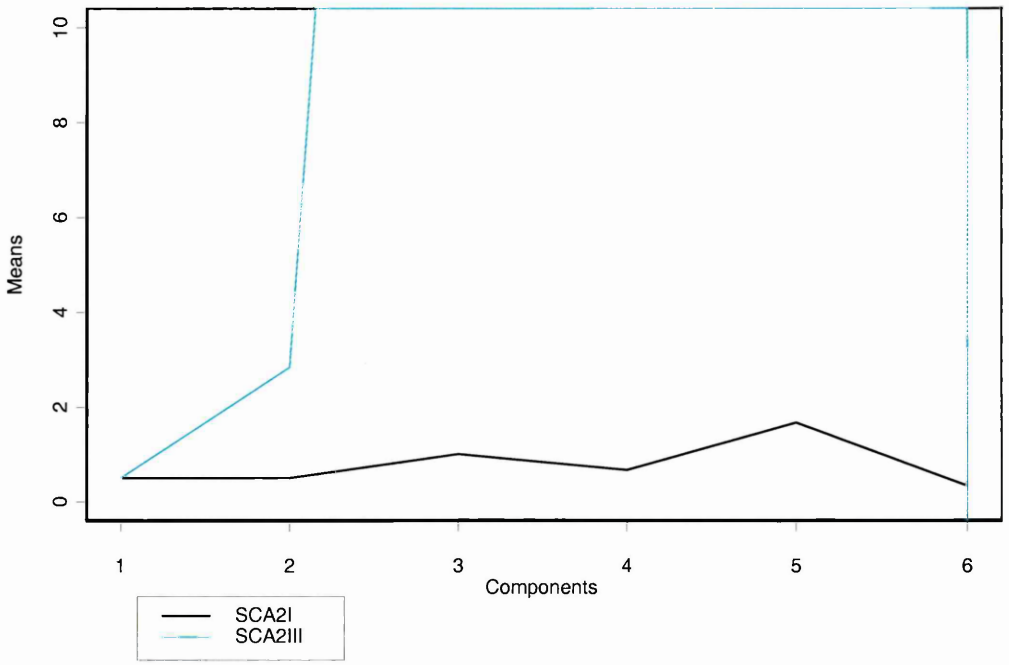

Figure 9.7: The simplicity of components produced by SCA2I and SCA2III for complex block structure

Section 9.4.1. But for simple uniform structure, the components produced by SCA2III are more simple than those produced by SCA2I.

So, it is concluded that for complex structures, in general, the results of SCA2 using restrictions I are more accurate and more simple than those using restriction III.

9.5 Results applied to V_0 when $k = 1$ for large dimensional data

In this section, I will investigate the population results obtained by SCA2 using restrictions I and III for large dimensional data. In Chapter 8, it was shown that for block structure (simple or complex) and simple uniform structure the results obtained by SCA2 when $k = 1$ are less

accurate than those obtained by SCA2 when $k = 0$. This is because SCA2 can get simple block and uniform structures back exactly when $k = 0$ but not when $k = 1$. In contrast, the results obtained by SCA2 for complex uniform structure and intermediate structures (simple and complex) are more accurate when $k = 1$ than when $k = 0$. Furthermore the conclusions from the sample simulation results generally are the same as that when applied to population variance covariance matrix V_0. So in this section, I just consider applying SCA2 using restrictions I and III with $k = 1$ to population covariance matrices V_0 for the 8 dimensional and 10 dimensional simple intermediate structures

9.5.1 Simple intermediate structure for 8 and 10 dimensional data

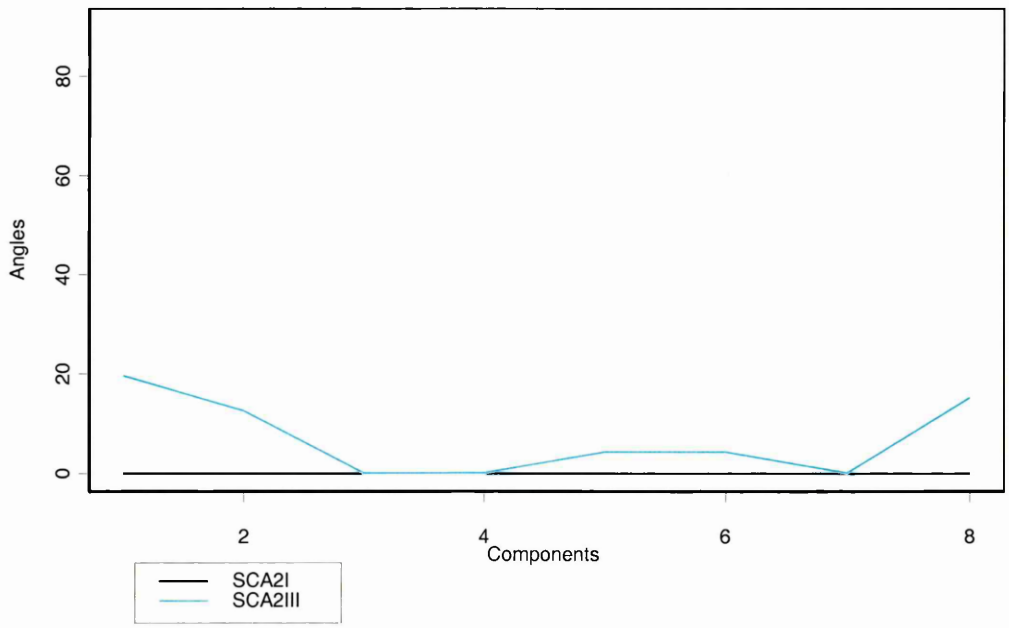

Figure 9.8: The accuracy of components produced by SCA2I and SCA2III for 8 dimensional simple intermediate structure

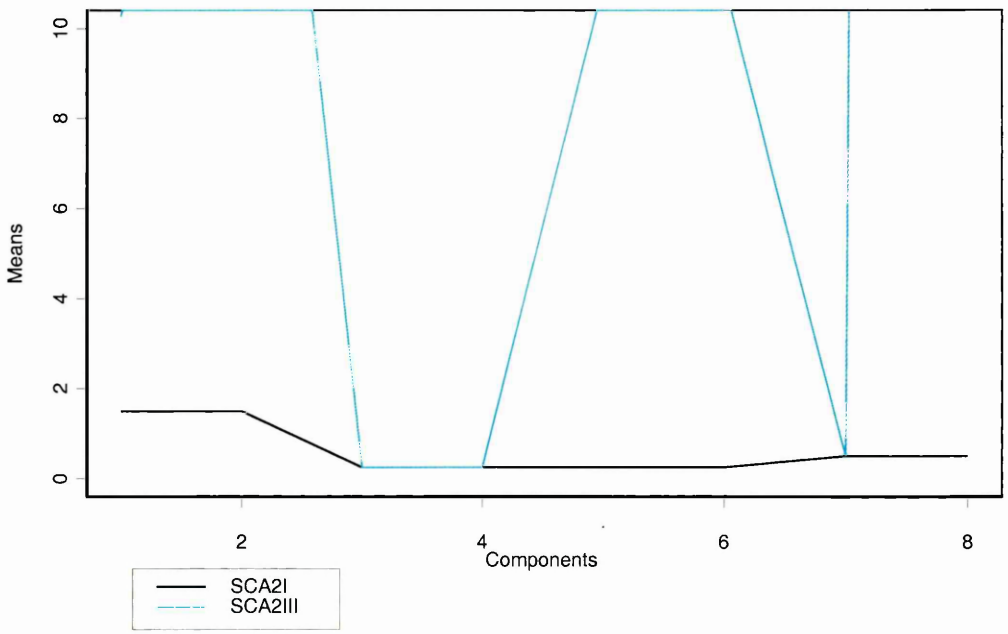

Figure 9.9: The simplicity of components produced by SCA2I and SCA2III for 8 dimensional simple intermediate structure

For the 8 dimensional simple intermediate structure, when $k = 1$, SCA2I retrieves the simple intermediate structure (Figure 9.8). This is better than that achieved by SCA2I for 6 dimensional simple intermediate structure. The results obtained by SCA2III are less accurate than those obtained by SCA2I. The results obtained by SCA2I are simple (Figure 9.9), this is expected because SCA2I retrieves the simple intermediate structure. The results obtained by SCA2III are not as simple as those obtained by SCA2I. In fact, the first two components produced by SCA2III are complex. That is expected considering the results in Section 9.4 for the 6 dimensional data. SCA2I retrieves the simple intermediate structure, so SCA2I gets the best results for 8 dimensional simple intermediate structure.

For the 10 dimensional simple intermediate structure, the results obtained by SCA2III are better than those obtained by SCA2I for the first two components but neither are very good.

As was said in Section 8.4.1, SCA5 and SCA6 using restriction I retrieve the 10 dimensional simple intermediate structure when $k = 1$ but not using restriction III. So restriction I is the best choice for 10 dimensional simple intermediate structure.

All in all, the results for large dimensional data obtained by SCA2III generally are neither as simple as nor always more accurate than those found by SCA2I. Again the same conclusions as for 6 dimensional data are obtained. In general, SCA2 using restriction I gets the best results for large dimensional data compared to the results obtained by SCA2 using restriction III.

9.6 Discussion and conclusion

In Chapter 9, three different restrictions on the directions considered for each simplicity preserving transformation were investigated. All the three restrictions are the same when $k = 0$, and restrictions I and II are always the same when $k = 1$ (Sections 9.2 and 9.3). SCA2 is one of the best SCA methods and the best results generally are obtained by SCA2 when $k = 0$ or 1. So in Chapter 9, only one method, SCA2, using restrictions I and III when $k = 1$ was investigated. It was shown in Sections 9.2 and 9.3 that when $k = 1$, the number of directions available for any individual SCA method using restriction III is more than those using restriction I. However, the results obtained by SCA2 using restriction III are not always more accurate than those using restriction I (for example, the population results for simple block structure in Section 9.4.1). The greater number of directions available for restriction III at each step does not guarantee that the maximal improvement in variance using restriction III is more than the maximal improvement in variance using restriction I because sometimes the best maximal improvement in variance might be obtained in a direction available using restriction I but not using restriction III. Even if the maximal improvement in variance of SCA2 using restriction III at each step is more than that using restriction I, this does not guarantee more accurate results overall (an example of this was given in Section 8.2.1) because this does not guarantee the maximal improvement in variance overall.

The best results obtained by the SCA methods are defined as the most accurate results among the simple results. When $k = 0$ the results obtained by SCA methods using any restrictions

are the same, and they retrieve the simple block and uniform structures. So, if the eigenvectors structure is block structure or uniform structure, just use $k = 0$ and hence the exact choice of restriction is irrelevant. SCA methods using restriction I obtained the best results for the intermediate structures. So if one restriction from these two restrictions has to be chosen for these data, restriction I is the best choice for the data introduced in Sections 4.3 and 6.2.

If the eigenvectors matrix has an intermediate structure, the best choice of k is 1 (Section 8.5), the best restriction is based on the form of the eigenvectors structure. Restriction I is the best choice for the intermediate structures introduced in Sections 4.3 and 6.2 because the loadings of the transformation matrix are similar to the loadings of the simple intermediate structures for two dimensional data (Sections 4.7 and 8.5) when $k = 1$. However, if the eigenvector matrix is

$$A = \begin{pmatrix} 1 & -3 & 1 & 1 & 1 & 0 \\ 1 & -3 & 1 & 1 & -1 & 0 \\ 1 & -3 & -2 & -2 & 0 & 0 \\ 3 & 1 & -1 & 1 & 0 & 1 \\ 3 & 1 & -1 & 1 & 0 & -1 \\ 3 & 1 & 2 & -2 & 0 & 0 \end{pmatrix},$$

SCA2 using restriction III when $k = 1$ gets this structure back exactly. So, in this case, restriction III is the best choice.

In order to get good results for different structures, new restrictions can be introduced, for example, by just changing 2^k the term in restriction I to 5^k, 6^k, 7^k etc according to the structure of the eigenvector matrix. The restrictions using 4^k, 8^k and 9^k do not need to be introduced because $4^k = 2^{2k}$, $8^k = 2^{3k}$ and $9^k = 3^{2k}$. The new restrictions using large integers (say, 11^k) are also not necessary. This is because in this case the components obtained by SCA methods generally are not simple. So the structure of the data (this is obtained by PCA) is a major factor to choose a suitable restriction rather than the number of the directions available at each iteration for SCA methods to get good results.

In practice, the components of the data sets generally are a combination of components of complex block, uniform and intermediate structures. So it is difficult to decide the structure of the eigenvectors matrix. For example, the PCA for the Jeffers' pitprop data (Table 1.9, Section

1.3.4). If the loadings less than 0.3 are set to 0, the eigenvectors have a block structure. So the best choice of k is 0. The best results are obtained by SCA2 when $k = 0$ (Section 3.5.4). If the loadings less than 0.1 are set to zero, the eigenvectors have an intermediate structure. The loadings in the eigenvector matrix generally are one times, two times, three times or four times of 0.1. So the best results should be obtained by SCA2, SCA5 or SCA6 with $k = 1$ or 2 using restrictions I or $k = 1$ using restrictions III. The results of SCA5 and SCA6 with $k = 1$ and 2 using restriction I and with $k = 1$ using restriction III are complex. The first two angles of SCA2 with $k = 1$ and 2 using restriction I and with $k = 1$ using restriction III are the same, and the other two angles are very similar. These results are less accurate than the results obtained by SCA2 when $k = 0$. So $k = 0$ is the best choice for the pitprop data.

The analysis in Chapter 9 did not consider all cases. The conclusions obtained by SCA2 can be extended to other variants of the SCA algorithm and applied to other data. In order to compare different restrictions, it is enough to just compare the results of SCA2, SCA5 and SCA6 using restrictions I and III when $k = 1$ unless the restrictions are different when $k = 0$.

Chapter 10

Combined results

10.1 Introduction

The results described in Chapters 4 to 9 showed that different SCA methods obtain the most accurate results for different components. For example, SCA2 got the most accurate results for components 1, 3, 4, 5 and 6 for the 6 dimensional complex uniform structure (Sections 4.6.2), whereas SCA5 got the most accurate results for components 1 and 2. The population results obtained by an individual SCA method using different values of k and different restrictions might yield the most accurate results for different components for a same data set. For example, for the 6 dimensional complex block structure (Section 9.4.2), SCA2 using restriction I got the most accurate results for components 1, 2 and 6, but SCA2 using restriction III got the most accurate results for components 3, 4, 5 and 6.

If the most accurate results are not simple, the most accurate results are not useful to enhance the interpretation for corresponding PCs. For example, the results obtained by SCA2 using restriction III for components 3, 4, 5 and 6 are very accurate but they are complex. So in this chapter, I only consider the most accurate results within all the simple results. These results are called the best results. Can the best results for every component be combined to form new results? The answer is "yes". This new approach is called the combining approach. The results obtained by combining approaches are called combined results.

Section 10.2 gives a description of the combining approach. Section 10.3 introduces ways to measure the variance explained by the components of the combined results and ways to evaluate the dimension-reducing power of the components of the combined results compared to principal components. In order to make the concept of combined results more clear, 3 kinds of combined result will be given in the following 3 sections. Section 10.4 gives the combined results using different SCA methods when $k = 0$. Section 10.5 discusses the combined results using different SCA methods with restrictions I and III when $k = 1$. Section 10.6 illustrates the combined results using different SCA methods with restrictions I and III and $k = 0$ and 1. The last section, Section 10.7, is the conclusion and discussion.

10.2 Description of the combining approach

It is known that, for a data set, the best results for different components may be obtained by different SCA methods using different values of k and different restrictions (Section 4.6.2). The combining approach combines the best results for every component from different SCA methods using different restrictions and different values of k. For a given component, the best component produced by SCA methods must be simple and the angle between the best component and corresponding PC must be the smallest out of the simple components. If all the results obtained by different SCA methods using different values of k and different restrictions are not simple for a given component, it means that SCA methods fail for this given component.

So the combined results have following two properties.

i) All the components of combined results are simple.

ii) The angle between a given component for combined results and corresponding principal component is the smallest within all the simple components for this component.

In general, the combined results are not orthogonal because the combined results might come from different SCA methods using different values of k and different restrictions.

Take the combined results obtained by applying SCA2 to the RI data (Section 1.3) using restrictions I and III when $k = 1$ as an example. As before, the best result of a component must be simple and have the smallest angle within all the simple components for this component. All

Restrictions	Component	variance	Angle	Directions			
I	1	4.68	10°	1	1	2	2
	2	2.15	2°	-1	1	-2	2
	3	0.92	10°	2	2	-1	-1
	4	0.54	2°	2	-2	-1	1
III	1	4.80	4°	2	2	3	3
	2	2.12	8°	-2	2	-3	3
	3	0.80	4°	3	3	-2	-2
	4	0.57	8°	3	-3	-2	2

Table 10.1: Simple components for the RI data using SCA2 when $k = 1$

Component	Restriction	variance	Angle	Directions			
1	III	4.80	4°	2	2	3	3
2	I	2.15	2°	-1	1	-2	2
3	III	0.80	4°	3	3	-2	-2
4	I	0.54	2°	2	-2	-1	1

Table 10.2: Combined results of RI data using SCA2 under different restrictions

the components obtained by SCA2 using restrictions I and III are simple (Table 10.1), so the best results are those with the smallest angles. The best results for components 1 and 3 are obtained using restriction III. Similarly, the best results for components 2 and 4 are obtained using restriction I. The combined results for the RI data using SCA2 when $k = 1$ are given in Table 10.2. The combined results are better than any just using SCA2 under any individual restriction when $k = 1$. The combined results here are also orthogonal, although generally they are not.

By the definition of combined results, the combined results could come from different SCA methods when $k = 0$ (case 1), from different SCA methods using restriction I or III when $k = 1$ (case 2), or from different SCA methods using restriction I or III when $k = 0$ or 1 (case 3) etc. For the same data set, the combined results in case 3 are the best overall because the combined results in case 3 are the combined results from cases 1 and 2. In order to make the concept of combined results more clear, in this chapter I mostly just consider the combined results of the population results for the RI data, the sparrow data, the employment in Europe data and the Jeffers' pitprop data. These four examples were introduced in Section 1.3. Also for case 3, I consider the combined results for the 6 dimensional data sets introduced in Section 4.3.

10.3 Ways to measure the variation explained and the dimension-reducing power of components

The column vectors of the loading matrix of the combined results, in general, are non-orthogonal and correlated, so the combined results provide redundant information. It is essential to know how to calculate the percentage of variance of the first m components of combined results explained.

Before starting reviewing the existing criteria, I quickly review some notation in Section 1.2. Suppose a linear dimension reduction technique is represented as $W = XC$, where X is a $n \times p$ matrix, n is the number of observations, p is the number of variables, without loss of generality, the column means of X are assumed to be 0. C is a $p \times m$ matrix, called the loading matrix and

$m \leq p$, γ_k is the k^{th} column of matrix C, C is a unit-norm loading matrix. The k^{th} component of this linear combination is then $W_k = X\gamma_k$. Let Σ be equal to the variance-covariance matrix of X.

The principal components are defined as follows. The variance covariance matrix Σ can be written as $\Sigma = A^T \Lambda A$, where $\Lambda = diag(\lambda_1, \ldots, \lambda_m)$, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$, λ_k is the k^{th} largest eigenvalue of Σ and the k^{th} column of A is the corresponding eigenvectors. The loading of $Z = XA$ are then the principal components, and A is called the loading matrix.

Suppose $var(W_k)$ is defined as the k^{th} element on the diagonal of matrix $C^T \Sigma C$. If W is uncorrelated, the total variance explained by the first m components is $\sum_{k=1}^{m} var(W_k)$. However, if W is correlated, the computed total variance by the first m components is not right because it does not handle correlation and non-orthogonality in a adequate way. It implies that the variance explained by one component is unaffected by that of another component, so $\sum_{k=1}^{m} var(W_k)$ might be greater than $\sum_{k=1}^{m} \lambda_k$. It is not true because principal components are orthogonal, uncorrelated, and extract the maximal variability (Jolliffe (2002b), Chapters 1 and 2). It was also proved by Okamoto (1969) and McCabe (1984) that principal components are optimal under different criteria that apply to uncorrelated or orthogonal components. So for a given m, the first m principal components are the optimal dimension-reducing system.

If W is correlated, a new criterion to compute the total variance by the first m components has been proposed by Gervini and Rousson (2004). The idea is that if a new component $W_k = X\gamma_k$ is added to a system of $(k-1)$ components, an indicator of the real contribution of W_k to the total variance of the system is the residual variance of the linear prediction of W_k given the first $k-1$ components. i.e. the variance explained by the k^{th} component is $var(W_k) - var(W_k | W_1, \ldots, W_{k-1})$. The total variance corrected for correlations are obtained by adding all these variances together. That is

$$K(C) = \sum_{k=1}^{m} (\gamma_k^T \Sigma \gamma_k - \gamma_k^T \Sigma C_{(k-1)} (C_{(k-1)}^T \Sigma C_{(k-1)})^{-1} C_{(k-1)}^T \Sigma \gamma_k)$$

where $C_k = (\gamma_1, \ldots, \gamma_k)$. The percentage of variance accounted by the first m components of W is

$$K(C) / \sum_{k=1}^{p} \lambda_k.$$

There are some other criteria for correlated components such as the criteria proposed by Chipman and Gu (2003) and Zou, Hastie and Tibshirani (2004). These criteria will not be considered here because they are difficult to apply.

To evaluate the dimension-reducing power of components compared to principal components, there are a lot of existing criteria.

First, consider the criterion connected with the variability maximization approach. Okamoto (1969) provided a criterion $GV(C)$ under the restriction of unit-norm loadings.

$$GV(C) = \left(\frac{det(C^T \Sigma C)}{\prod_{k=1}^{m} \lambda_k} \right)^{\frac{1}{m}}.$$

The m components with maximal variance are the ones that carry most of the information of the original data, while the other vary little about zero. The drawback of this criterion is that GV is invariant for SCA methods. For any SCA method, $det(C^T \Sigma C) = det(\Sigma)$ because C is a orthogonal matrix (in this formula,the length of each column of C is 1), this makes the performances of different SCA methods cannot be discriminated by GV. If the combined results are orthogonal, this also makes the performance of different combined results can not be discriminated by GV. So this criterion is not suitable for combined results. Two new criteria, CSV(C) and SCSV(C), have been proposed by Gervini and Rousson (2004). $CSV(C)$ is defined to be

$$CSV(C) = \frac{\sum_{k=1}^{m} (\gamma_k^T \Sigma \gamma_k - \gamma_k^T \Sigma C_{(k-1)} (C_{(k-1)}^T \Sigma C_{(k-1)})^{-1} C_{(k-1)}^T \Sigma \gamma_k)}{\sum_{k=1}^{m} \lambda_k}.$$

The first term of CSV(C) is equal to the variance of the first component divided by $\sum_{i=1}^{m} \lambda_k$. The k^{th} term of CSV(C) is the variance of the k^{th} component corrected by the first $(k-1)$ components divided by $\sum_{i=1}^{m} \lambda_k$. The numerator of CSV(C) is just the sum of variances of the components if the components are uncorrelated ($C_{(k-1)}^T \Sigma \gamma_k = 0$). Another alternative is to define a corrected sum of variances

$$SCSV(C) = \frac{\sum_{k=1}^{m} (\gamma_k^T \Sigma \gamma_k - \gamma_k^T \Sigma C_{-k} (C_{-k}^T \Sigma C_{-k})^{-1} C_{-k}^T \Sigma \gamma_k)}{\sum_{k=1}^{m} \lambda_k}.$$

where C_{-k} is the $p \times (m-1)$ matrix obtained after deleting the k^{th} column of C. The numerator of $SCSV(C)$ is the sum of the residual variances of the linear predictors of W_k given

the other $(m-1)$ components. If the components are uncorrelated, $CSV(C) = SCSV(C)$. $SCSV(C)$ measures the variance each components explains in a different way because in the second case, the variance of the k^{th} component is corrected by the other $(m-1)$ components. These two criteria are suitable criteria for the combined results.

There are many other criteria. I don't elaborate on these. More details can be found in Gervini and Rousson (2004).

Suppose the variance of the j^{th} simple component obtained by SCA method i $(i = 1, 2, 3, 4, 5, 6)$ using restriction l (l= I and III) is represented by μ_{ilj} ($j = 1, 2, \ldots, p$). The criterion

$$L(C) = \frac{\sum_{j=1}^{m} \mu_{ilj}}{\sum_{k=1}^{m} \lambda_k}.$$

is generally used to measure the percentage of variance accounted by the first m components of combined results, where i and l generally are different for different j because for different component j of the combined results, the best results might come from different SCA method using different restrictions. So the percentage of loss in total variance is $1 - L(C)$. This criterion generally overestimates or underestimates the total variance explained by the first m components when the non-orthogonal and correlated components of combined results are used. However, this formulation is easy to use and in practice the approximation is sometimes acceptable.

So in the following examples, if all the components are orthogonal, though they are correlated in general, $L(C)$ will be used. Otherwise $CSV(C)$ will be used.

10.4 Examples of using different SCA methods, $k = 0$

In this section the combined results applied to all the four data sets introduced in Section 1.3, the RI data, the sparrow data, the employment in Europe data and the Jeffers' pitprop data will be investigated.

When $k = 0$, restrictions I and III (Section 9.2) are identical. So the combined results just come from different SCA methods.

Comp.	Method	var.	PCA var.	Ang.	Directions
1	SCA1, SCA5 and SCA6	4.73	4.81	8°	1 1 1 1
2	SCA1, SCA5 and SCA6	1.97	2.15	19°	-1 1 -1 1
3	SCA1, SCA5 and SCA6	0.87	0.79	8°	1 1 -1 -1
4	SCA1, SCA5 and SCA6	0.72	0.54	19°	-1 1 1 -1

Table 10.3: Combined results for the RI data using different methods, $k = 0$

10.4.1 RI data

First, consider the combined results of applying different SCA methods to the RI data (Section 1.3.1). In this example, SCA1, SCA5 and SCA6 all gave the same results, and the results from SCA1, SCA5 and SCA6 are the best for all the components. So the combined results (Table 10.3) are the same as those just using SCA1, SCA5 or SCA6. The combined results come from the same SCA method, so the components of the combined results here are necessarily orthogonal. The cumulative variance explained by the first m ($m \leq 4$) components is 57%, 81%, 92% and 100% respectively. If the first component is retained, $L(C) = 98\%$, the combined results lose 2% dimension-reducing power compared to principal components. If the first two components are retained, $L(C) = 96\%$, the combined results lose 4% dimension-reducing power compared to principal components. If the first three components are retained, $L(C) = 98\%$, the combined results lose 2% dimension-reducing power compared to principal components. If all the components are retained, $L(C) = 100\%$, the combined results have no loss of the dimension-reducing power and no loss of dimension compared to principal components, this is expected because the transformation matrix of SCA methods have the same trace as that of original variance-covariance matrix.

10.4.2 Sparrow data

Next, consider the combined results for the sparrow data (Section 1.3.2) using different SCA methods. The combined results (Table 10.4) come from the same method SCA5 or SCA6. The

Comp.	Method	var.	PCA var.	Ang.	Directions				
1	SCA5 or SCA6	3.61	3.62	3°	1	1	1	1	1
2	SCA5 or SCA6	0.52	0.53	14°	0	1	1	0	-2
3	SCA5 or SCA6	0.37	0.39	19°	1	1	-1	-1	0
4	SCA5 or SCA6	0.31	0.30	11°	1	-1	1	-1	0
5	SCA5 or SCA6	0.19	0.16	18°	3	-2	-2	3	-2

Table 10.4: Combined results for the sparrow data using different methods, $k = 0$

Comp.	Method	var.	PCA var.	angle	Directions								
1	SCA2	3.36	3.49	12°	1	0	-1	-1	-1	-1	0	-1	-1
2	SCA5	1.80	2.13	35°	0	1	0	0	0	-1	-1	0	0
3	SCA5	1.07	1.10	-40°	-1	-1	0	-2	0	0	-1	1	0

Table 10.5: Combined results for the employment in Europe data using different methods, $k = 0$

components are therefore orthogonal and the total variance explained by the first m ($m \leq 5$) components in turn is 72%, 82%, 89%, 95% and 100%. If the first component and first two components are retained, $L(C)$ is almost 100%, the combined results have no lost of dimension-reducing power compared to principal components. If the first three components or first four components are retained, $L(C) = 99\%$, the combined results lose 1% of dimension-reducing power compared to principal components. If all the components are retained, $L(C)$ must be 100% because this combined results come from SCA6, the orthogonality of simple components makes the PCA and SCA account for same amount of variance if all the components are retained.

10.4.3 Employment in Europe data

Third, consider the combined results for the employment in Europe data (Section 1.3.3) using different SCA methods. As in Section 1.3.3, only the first three components are retained. The

Comp.	Meth.	var.	PCA V.	ang.	Directions												
1	SCA2	3.92	4.22	18°	1	1	0	0	0	1	1	1	1	1	0	0	0
2	SCA2	1.97	2.38	43°	1	1	1	1	-1	-1	-1	0	0	0	0	0	1
3	SCA5	1.81	1.88	-34°	2	2	-9	-9	-5	-11	-9	8	1	8	0	-7	3
4	SCA6	1.07	1.11	28°	-1	-1	-1	-1	1	1	1	5	-3	-7	16	-3	-1

Table 10.6: Combined results for the Jeffers' pitprop data using different methods, $k = 0$

combined results (Table 10.5) come from SCA2 and SCA5, components 1 and 2 are not orthogonal, the corrected variances are 3.36, 1.64 and 0.95 respectively, the total variance explained by the first m ($m \leq 3$) components is 37%, 54% and 65% respectively. If the first component is retained, $CSV(C) = 96\%$, the combined results lose 4% dimension-reducing power compared to principal components. If the first two components are retained, $CSV(C) = 89\%$, the combined results lose 11% dimension-reducing power compared to principal components. If the first three components are retained, $CSV(C) = 88\%$, the combined results lose 12% dimension-reducing power compared to principal components.

10.4.4 Jeffers' pitprop data

Finally, consider the combined results for Jeffers' pitprop data (Section 1.3.4) using different SCA methods. As in Section 1.3.4, only the first four components are retained. The combined results (Table 10.6) for this data come from SCA2, SCA5 and SCA6. Components 1 and 3, components 1 and 4, 2 and 3 , components 2 and 4, components 3 and 4 are not orthogonal. Although the combined results for components 3 and 4 are more accurate than those of SCA2 when $k = 0$, components 3 and 4 of the combined results become more difficult to interpret compared to the components 3 and 4 produced by SCA2 when $k = 0$ (Section 3.5.4, Table 3.10). So I don't compute the corrected variance by CSV(C). For this data, I think the results produced by SCA2 when $k = 0$ got the best tradeoff between accuracy and simplicity. So I will not investigate the combined results for Jeffers' pitprop data in other cases.

Comp.	Method	Res.	var.	PCA var.	Ang.	Directions			
1	SCA6	I	4.81	4.81	1°	3	3	4	4
2	SCA2 and SCA3	I	2.15	2.15	2°	-1	1	-2	2
3	SCA6	I	0.79	0.79	1°	-4	-4	3	3
4	SCA2 and SCA3	I	0.54	0.54	2°	2	-2	-1	1

Table 10.7: Combined results for the RI data under restrictions I and III using different methods, $k = 1$

In fact from Tables 10.3, 10.4 and 10.5 it appears that the combining approach has not been successful. In all three cases using a suitable individual method would have the same results as the combining approach. However, the combined results of Jeffers' pitprop data in Table 10.6 are more accurate than those using any individual SCA method. So in next section, the combined results using different SCA methods with different restrictions when $k = 1$ will be discussed.

10.5 Examples of using different SCA methods and different restrictions, $k = 1$

In this section, the combined results applied to the first 3 data sets as in Section 10.4 will be considered, but this time the combined results come from different SCA methods as well as different restrictions when $k = 1$.

10.5.1 RI data

The first example is the combined results for the RI data using different SCA methods with restrictions I and III when $k = 1$. The components of the combined results for the RI data are orthogonal (Table 10.7). The first m ($m \leq 4$) components account for 58%, 84%, 94% and 100% variance respectively. Comparing the results in Table 10.7 with those in Table 10.2, the

Comp.	Method	Res.	var.	PCA var.	Ang.	Directions				
1	SCA2	I and III	3.61	3.62	3°	1	1	1	1	1
2	SCA6	I	0.53	0.53	8°	0	3	4	1	-8
3	SCA6	III	0.35	0.39	27°	2	3	-2	-3	0
4	SCA6	I and III	0.31	0.30	11°	-1	1	-1	1	0
5	SCA2	I	0.20	0.16	23°	3	-7	-1	5	0

Table 10.8: Combined results for the sparrow data using different methods and restrictions I and III, $k = 1$

results in Table 10.7 are better than those obtained by SCA2 in Table 10.2. This is because the results here are the best results amongst all the SCA methods. The results here are also much better than the results in Table 10.3. This is because the eigenvector matrix of PCA for the RI data is similar to the intermediate structure (Section 9.6). However the difference is not too much, the gain is not worth the extra computation involved. The corresponding eigenvalues of these combined results are the same to a 2 decimal places as those of principal component for the RI data, so no matter how many components are retained, $L(C)$ are always 100%. So these combined results are very good.

10.5.2 Sparrow data

The second example is the combined results for the sparrow data using different SCA methods with different restrictions when $k = 1$. These combined results are better than those just using any individual method (Table 10.8). This is expected because all the SCA methods are considered here. The angles here are not smaller than the angles in Table 10.4 except the second angle. Components 2 and 3, 2 and 5, 3 and 5, 4 and 5 are not orthogonal, all other pairs of components are orthogonal with each other. The corrected variances obtained by CSV(C) are 3.61, 0.53, 0.34, 0.30, 0.10 respectively. The corrected variance of each component is very similar to the variance of each component obtained by L(C) except the last component. So the

Comp.	Method	Res.	var.	PCA var.	ang.	Directions								
1	SCA5	I	3.36	3.49	12°	-7	-1	4	3	4	4	1	3	3
2	SCA2	III	1.93	2.13	31°	2	3	2	0	2	-2	-3	-2	0
3	SCA2	I	1.01	1.10	87°	-2	0	-3	0	-3	2	0	2	0

Table 10.9: Combined results for the employment in Europe data using different methods under restrictions I and III, $k = 1$

conclusions don't change much. So the first m ($m \leq 5$) components of these combined results explain 72%, 83%, 91%, 96% and 98% of the variance respectively.

10.5.3 Employment in Europe data

The third example is the combined results of the employment in Europe data using different SCA methods under different restrictions when $k = 1$. For this example (Table 10.9), components 1 and 2, components 1 and 3, 2 and 3 are not orthogonal. In other words each pair of combined results are not orthogonal. The first angle here is the same as that in Table 10.5. The second angle here is smaller than the angle in Table 10.5 and the third angle bigger than the angle in Table 10.5. By CSV(C), the corrected variances are 3.36, 0.90 and 0.30 respectively. The corrected variance of the second components reduces to 0.90 from 1.93 and the corrected variance of the third component reduces to 0.30 from 1.10. So formulation $L(C)$ overestimates the variance accounted by the second and third components of combined results. The first m ($m \leq 3$) components of the combined results explain 37%, 47% and 50% the variance respectively. The combined results are expected to be at least as accurate as that using any individual method, because the results are the combined results using different SCA methods and different restrictions when $k = 1$. If the first three components are retained, $CSV(C) = 68\%$, the combined results lose 32% dimension-reducing power compared to principal components. Too much variance is lost by this combined results. So these combined results are not good. The extra computation does not seem worth this little improvement and the loss of total variance

explained by the first three components.

For the RI data, the sparrow data and the employment in Europe data, the combined results obtained by SCA methods using different restrictions when $k = 1$ are improved compared to those using any individual SCA method. Furthermore the combined results for the RI data and the sparrow data are very good. So in the next section, I will consider the combined results of SCA methods using different restrictions and $k = 0$ and 1.

10.6 Examples of using different SCA methods and different restrictions when $k = 0$ and 1

In this section, I will consider the combined results from all the SCA methods when $k = 0$ and all the SCA methods using restrictions I and III when $k = 1$. I am going to investigate the combined results applied to the same data sets used in Section 10.5 and the 6 dimensional data used in Section 4.3.

Section 10.4 gave the combined results using different SCA methods when $k = 0$ for the RI data, the sparrow data and the employment in Europe data. Section 10.5 gave the combined results using different SCA methods and restrictions I and III when $k = 1$ for these three data sets. So the combined results for these three data sets here are the combined results in Sections 10.4 and 10.5.

10.6.1 RI data

First consider the combined results using different SCA methods, different restrictions and $k = 0$ and 1 for the RI data. The simple components of combined results are orthogonal (Table 10.10), the first m ($m \leq 4$) components account for 58%, 84%, 94% and 100% variance respectively. The combined results here are the same as the combined results (Table 10.6) obtained by all the SCA methods using restrictions I and III when $k = 1$.

Comp.	Method	Res.	k	var.	PCA var.	Ang.	Directions			
1	SCA6	I	1	4.81	4.81	1°	3	3	4	4
2	SCA2 and SCA3	I	1	2.15	2.15	2°	-1	1	-2	2
3	SCA6	I	1	0.79	0.79	1°	-4	-4	3	3
4	SCA2 and SCA3	I	1	0.54	0.54	2°	2	-2	-1	1

Table 10.10: Combined results for the RI data using different SCA methods, different restrictions and $k = 0$ and 1

Comp.	Method	Res.	k	var.	PCA var.	Ang.	Directions				
1	SCA2	all	0, 1	3.61	3.62	3°	1	1	1	1	1
2	SCA6	I	1	0.53	0.53	8°	0	3	4	1	-8
3	SCA6	all	0	0.37	0.39	19°	1	1	-1	-1	0
4	SCA6	all	0, 1	0.31	0.30	11°	-1	1	-1	1	0
5	SCA6	all	0	0.19	0.16	18°	3	-2	-2	3	-2

Table 10.11: Combined results for the sparrow data using different SCA methods, different restrictions and $k = 0$ and 1

Comp.	Method	Res.	k	var.	PCA var.	ang.	Directions
1	SCA4	all	0	3.36	3.49	12°	1 0 -1 -1 -1 -1 0 -1 -1
2	SCA2	III	1	1.93	2.13	31°	2 3 2 0 2 -2 -3 -2 0
3	SCA2	all	0	1.07	1.10	43°	-1 -1 0 -6 2 1 -1 1 1

Table 10.12: Combined results for the employment in Europe data using different SCA methods, different restrictions and $k = 0$ and 1

10.6.2 Sparrow data

Next consider the combined results (Table 10.11) for the sparrow data. Components 2 and 3, 2 and 5 are not orthogonal, other components are orthogonal with each other. So CSV(C) is used. The corrected variances using CSV(C) are 3.61, 0.53, 0.37, 0.30 and 0.17 respectively. The corrected variances obtained by CSV(C) are very similar to the variances of PCA. So CSV(C) is almost 100% despite how many components are retained. The first m ($m \leq 5$) components account for 72%, 83%, 90%, 96% and 100% variance respectively. The variance for each component of combined results obtained by CSV(C) is very similar to the variance by L(C). The combined results given in Table 10.11 are better than that in Tables 10.4 and 10.8. This is expected, these results are at least as accurate as the better results between Tables 10.4 and 10.8, because these results are the combined results in Tables 10.4 and 10.8.

10.6.3 Employment in Europe

Then consider the combined results for the employment in Europe data. Components 1 and 2, 2 and 3 are not orthogonal (Table 10.12). In this example, using CSV(C), the corrected variance each component explained is 3.36, 0.91 and 0.85 respectively. The first m ($m \leq 3$) components of combined results explain 37%, 47% and 56% of the variance respectively. So the variances for the second and third components are overestimated (1.93 vs 0.91, 1.07 vs 0.85) by $L(C)$. If the first three components are retained, $CSV(C) = 76\%$, this means the combined results have lost 24% reducing-dimension power compared to principal components. These combined results

Comp.	Method	Res.	k	var.	PCA var.	Ang.	Directions					
1	SCA2 and SCA4	I	1	12.00	12	0°	1	1	1	2	2	2
2	SCA2	I	1	8.00	8	0°	2	2	2	-1	-1	-1
3	SCA2	I	1	6.00	6	0°	2	-1	-1	-1	-1	2
4	SCA5	all	0	4.15	4	23°	-2	13	-11	1	-13	12
5	SCA5	all	0	1.98	2	9°	-7	-7	14	4	-11	7
6	SCA2	all	0	1.14	1	22°	2	-1	-1	7	-5	-2

Table 10.13: Combined results for the 6 dimensional simple intermediate structure using different SCA methods, different restrictions and $k = 0$ and 1

are more accurate than the combined results in Tables 10.5 and 10.9. This is expected because these combined results combine the best components in Tables 10.5 and 10.9.

10.6.4 Simple structures- 6 dimensional data

Next, consider the combined results for the 6 dimensional data introduced in Section 4.3. For 6 dimensional simple block structure, SCA2, SCA5 and SCA6 retrieve the simple block structure when $k = 0$. This results can not be improved, the combined results are the same as those of SCA2, SCA5 and SCA6. SCA2 when $k = 0$ retrieves the simple uniform structure. So the combined results are the same as those of SCA2 when $k = 0$. For the simple intermediate structure, the combined results using different methods, different restrictions and different k worth doing because the SCA methods do not retrieve the simple intermediate structure. The combined results for simple intermediate structure are given in Table 10.13. Components 4 and 6, components 5 and 6 are not orthogonal (Table 10.13). In this example for component 6, the component with the smallest angle is not simple, so the component with the second smallest angle is chosen as component 6 of the combined results. As expected, the combined results are better than the results using any individual SCA method. The corrected variances of combined results by the CSV(C) are 12, 8, 6, 3.41, 1.92 and 0.94 respectively. Only the variances of

Comp.	Method	Res.	k	var.	PCA var.	Ang.	Directions					
1	SCA2	I	1	11.96	12	3°	1	1	1	2	2	2
2	SCA2	I	1	7.97	8	4°	2	2	2	-1	-1	-1
3	SCA2	I	1	5.97	6	6°	2	-1	-1	-1	-1	2
4	SCA2	all	0	3.52	4	27°	-1	1	0	0	-1	1
5	SCA2	all	0	2.33	2	30°	1	3	-4	0	1	-1
6	SCA2	all	0	1.15	1	20°	2	-1	-1	7	-5	-2

Table 10.14: Combined results for the 6 dimensional complex intermediate structure using different SCA methods, different restrictions and $k = 0$ and 1

the last three components are reduced because components 4 and 6, components 5 and 6 are not orthogonal. The first m ($m \leq 6$) components of the combined results account for 36%, 60%, 78%, 88%, 94% and 98% variance respectively. If the first three components are retained, $CSV(C) = 100\%$. If all the six components are retained, $CSV(C) = 98\%$, only 2% variance is lost. This means the combined results are very good.

10.6.5 Complex structures-6 dimensional data

For the complex block structure, the combined results from different SCA methods, different values of k and different restrictions are the same as the results obtained by SCA2, SCA5 and SCA6 when $k = 0$. For the complex uniform structure, the combined results from different methods, different values of k and different restrictions are the same as the results obtained by SCA2 when $k = 0$. This is because SCA2 (for complex block structure including SCA5 and SCA6) when $k = 0$ got the corresponding simple block and uniform structures back exactly. So generally the results of SCA2 are the best results for the complex block and uniform structures because the components of the simple structures generally are the closest simple approximation of corresponding complex structures. For the complex intermediate structure, the combined results from different methods, different k and different restrictions are given in Table 10.14.

All the components (Table 10.14) of combined results are orthogonal. In this example, the best results of components 4, 5 and 6 are not the smallest angles, because the results with the smallest angles are not simple. All the components here are orthogonal, the first m ($m \leq 6$) components of combined results account for 36%, 60%, 78%, 89%, 96% and 100% variance respectively. For this structure, SCA2 just need to be considered. The combined results are better than that just using SCA2 when $k = 0$ or 1. If the first one, first two, first three and all six components are retained, $L(C) = 100\%$, this means the combined results have no loss reducing-dimension power compared to principal components. If the first four components are retained, $L(C) = 98\%$. If the first five components are retained, $L(C) = 99\%$.

10.7 Discussion and conclusion

The combined results are at least as accurate as the best results obtained by using just any individual SCA method. So if more accurate results are needed, combined results are a good choice. However, generally, the combined results are not orthogonal.

The combined results can come from any SCA method using any restriction and any values of k. But for the examples in Chapter 10, generally, the best result of each component is obtained by SCA2, SCA5 or SCA6 using restriction I when $k = 0$ or 1. This is expected because SCA2, SCA5 and SCA6 are the best SCA methods, and SCA2, SCA5 and SCA6 using restriction I generally get the best results for the data in Sections 1.3, 4.3 and 6.2 when $k = 0$ or 1. For the data used in practice, it is generally enough to only consider the results obtained by SCA2, SCA5 and SCA6 using restrictions I and III and k is equal to 0 and 1.

There might be some problems for combined results. The i^{th} component of combined results might not be the best result of all the i^th components of the results obtained by SCA methods. For example, the third component of the combined results might be the second component obtained by an individual SCA method. Furthermore, the order of the component of combined results might not be displayed in the right order. For example, the variance corresponding to the second component of combined results might be less than the variance corresponding to the third component of combined results. Finally, it is possible for the dimensionality of the

combined results to be less than the dimensionality of the corresponding simple components. But I have not met these cases in practice. So generally the combining approach works well in practice.

Chapter 11

Hybrid approach

11.1 Introduction

It was shown in Chapters 4 to 8 that individual SCA methods are not ideal because, sometimes for the data given in Sections 4.3 and 6.2, individual SCA methods didn't retrieve the simple structure. Individual SCA methods also sometimes did not retrieve the corresponding simple structure of the complex structure either because generally the simple structure is the closest simple approximation of corresponding complex structure by construction. Furthermore sometimes the most accurate results obtained by SCA methods are not simple enough to interpret. If different SCA methods with different criteria and different values of k are used in different iterations, the results of SCA methods may hopefully be improved compared with individual SCA methods. And hybrid approaches are also expected to get the best tradeoff between accuracy and simplicity. Such an approach using different SCA methods in different iterations is called a hybrid approach. In contrast to combined results, the components produced by hybrid approaches are orthogonal because at every iteration the components transformed by SCA methods remain orthogonal.

Hybrid approaches use, in different iterations, different individual SCA methods with different criteria and different values of k. These individual SCA methods might all use single updating, all use multiple updating, or a mixture of single and multiple updating. However, at

any given iteration, the SCA method which is used is known in advance. For example, SCA4 might be used in the first iteration and SCA2 used in the remaining iterations. Alternatively SCA4 might be used in the first 2 iterations and SCA2 used in the remaining iterations. If SCA4 is used in the first 0 iterations and SCA2 in the remaining iterations, the hybrid approach is the same as SCA2. So an individual SCA method is a special case of hybrid approaches. Hybrid approaches can use more than two SCA methods. For example, SCA2 is used in the first iteration, SCA5 is used in the second iteration and SCA6 is used in the remaining iterations. But I will not explore this possibility in this chapter.

Recall from Chapters 4 to 9 that SCA2, SCA5 and SCA6 are the best individual SCA methods, and that the results of SCA5 and SCA6 become increasingly similar as the dimension of data increases. So it is only worth considering hybrid approaches which use a mixture of single updating with maximal improvement in variance (SCA2) and multiple updating with maximal improvement in variance (SCA5). Furthermore, the results of SCA2, SCA5 and SCA6 generally are accurate and simple when $k = 0$. Thus the hybrid approaches investigated in this chapter will always use the same criterion, maximal improvement in variance, and only use $k = 0$. So, the only thing that changes in the hybrid approach is the updating method. Two types of hybrid approaches will be discussed in this chapter. The first type uses multiple updating in the first few (say q) iterations (SCA5), and uses single updating in the remaining iterations (SCA2), which will be abbreviated as M^qS. The second type uses single updating in the first few iterations (q) and multiple updating in the remaining iterations, which will be abbreviated as S^qM.

The means of the components produced by hybrid approaches discussed in this chapter generally are between those of SCA2 and SCA5. This is because the hybrid approaches generally use multiple updating and single updating together, and the results using single updating generally are more simple than those using multiple updating (Section 4.7). So in the following examples I will not give more details of the simplicity of the components produced by hybrid approaches except for M^2S.

Section 11.2 illustrates the population results produced when hybrid approaches M^qS and S^qM with $q \geq 1$, SCA2 and SCA5 are applied to 6 dimensional data. Section 11.3 gives the

population results produced when hybrid approaches M^qS and S^qM with $q \geq 1$, SCA2 and SCA5 are applied to 8 dimensional data. Section 11.4 gives the population results produced when hybrid approaches M^qS and S^qM with $q \geq 1$, SCA2 and SCA5 are applied to real data. The last section, Section 11.5, is the discussion and conclusion.

11.2 Hybrid approaches for 6 dimensional data, $k = 0$

In this section, the population results obtained when hybrid approaches are applied to 6 dimensional data will be discussed. The eigenvectors matrices for 6 dimensional data used were given in Section 4.3. Two types of hybrid approaches will be implemented: using SCA5 first followed by SCA2, M^qS ($q \geq 1$), and using SCA2 first followed by SCA5, S^qM ($q \geq 1$). In particular, the results of M^2S, SCA2 and SCA5 will be compared in this section.

11.2.1 Simple block structure

Recall from Section 4.5.1 that when SCA2 and SCA5 are applied to the population variance covariance matrix V_0 with simple block structure, they retrieve the simple block structure. So for simple block structure, the results obtained by any hybrid approach can not be more accurate than those of SCA2 and SCA5.

Actually the two types of hybrid approaches, M^qS and S^qM with $q \geq 1$, also retrieve the simple block structure. Take hybrid approach M^1S as an example. Recall from Section 4.5.1, the results of SCA2 can be written as $P_1P_2P_3P_4P_5$, and the results of SCA5 as $P_1P_3P_2P_4P_5$, where P_j, $j = 1, \ldots, 5$ is the same as the transformation matrix of SCA2 in iteration j. P_3 and P_2 can be exchanged because they update different directions, leading to SCA2 and SCA5 giving the same results. M^1S took 4 iterations. In the first iteration, the transformation matrix of M^1S is the same as that of SCA5 by definition, i.e. P_1P_3. In iterations 2 to 4, the transformation matrices of this hybrid approach are P_2, P_4 and P_5 respectively. So this hybrid approach gets the same results as SCA5 and SCA2. The results of other hybrid approaches can be explained in a similar way. So, in this a special case all the hybrid approaches get the same results as SCA2 and SCA5. In general, even if SCA2 and SCA5 get the same results, it is possible for

some of the hybrid approaches to get different results because generally transformation matrices can not be exchanged.

11.2.2 Simple uniform structure

It is known that SCA2 gets the simple uniform structure back exactly (Section 4.5.2). In contrast, using SCA5 the simple uniform structure is not returned exactly. So for the 6 dimensional simple uniform structure any hybrid approach cannot get more accurate results than those obtained by SCA2 but there is room for a hybrid approach to do better than SCA5.

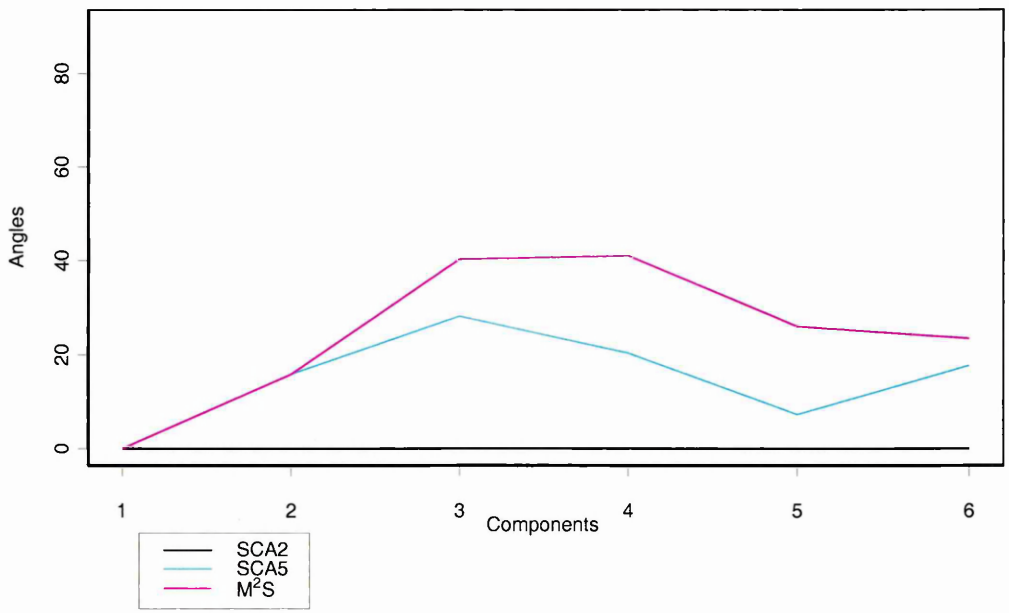

Figure 11.1: The angles obtained by M^2S, SCA2 and SCA5 for 6 dimensional simple uniform structure

The results obtained by the hybrid method M^2S (Figures 11.1, 11.2) are less accurate than those obtained by SCA2 and SCA5. The results of hybrid method M^2S are the same as those of SCA5 for the first two components. The components produced by SCA2 are more simple

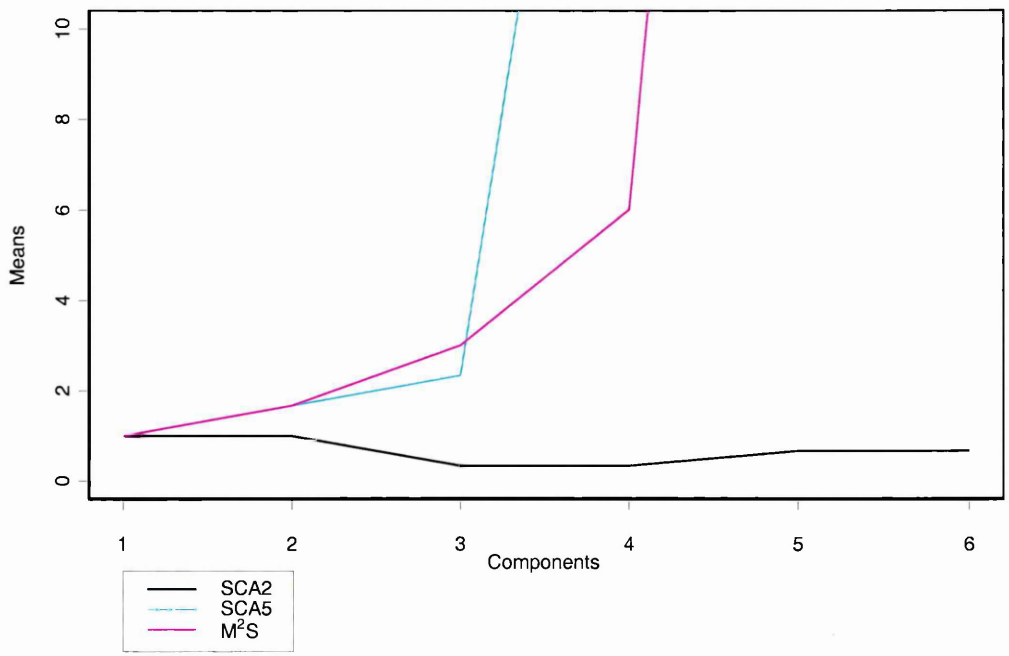

Figure 11.2: The simplicity of the components produced by M^2S, SCA2 and SCA5 for 6 dimensional simple uniform structure

than those produced by hybrid method M^2S and SCA5, whereas the means of the first three components produced by M^2S are more than those of SCA2 and SCA5. It is a little bit of a surprise. I suppose the results (both angles and means) obtained by hybrid approaches generally are between both those obtained by SCA2 and SCA5 because hybrid approaches use SCA2 and SCA5 in different iterations. Instead this shows that it is possible for the results of the hybrid method M^2S to be less accurate than both of those obtained by SCA5 and SCA2. One explanation is that hybrid method is equivalent to applying SCA2 to the new variance covariance matrix V produced after SCA5 is applied to V_0 for two iterations. This might lead to less or more accurate results compared to SCA5 and SCA2, this also might lead to less or more simple results compared to SCA5 and SCA2.

In order to investigate the full range of possible hybrid approaches which combine SCA2 and

SCA5 in different iterations, it is necessary first to check how many iterations SCA2 and SCA5 took. It is not necessary to consider larger values of q than the number of steps SCA2 or SCA5 took because then the first type hybrid approach is equivalent to SCA5 and the second type hybrid approach is equivalent to SCA2. SCA2 took 7 iterations, SCA5 took 8 iterations. So I just consider the first type hybrid approach M^qS with $q = 1, \ldots, 7$, the second type of hybrid approach S^qM with $q = 1, \ldots, 6$. Hybrid approaches use SCA2 and SCA5 in different iterations. So hopefully at least there is a hybrid approach which can get the best balance between accuracy and simplicity.

The results obtained by hybrid approach MS are the same as those obtained by hybrid method M^2S. The hybrid approaches M^qS, $q = 3, \ldots, 7$, get the same results as those of SCA5. This first type of hybrid approach M^qS is more likely to get similar results to SCA5 as q increases because the higher the q, the more iterations of M^qS ($q \geq 1$) are like SCA5 iterations.

Next consider the second type of hybrid approach, S^qM ($q \geq 1$). Hybrid approach SM gets the simple uniform structure back exactly. In other words, the results of SM are the same as those obtained by SCA2 and better than those obtained by SCA5. Also this hybrid approach only took 4 iterations, whereas SCA2 took 7 iterations, so hybrid approach SM is better than SCA2. The results obtained by hybrid method, S^2M, are less accurate than those of SCA2, but more accurate than those of SCA5 and hybrid method M^2S. All the other hybrid approaches, S^qM, $q = 3, \ldots, 6$, get the same results as those of SCA2. So, as expected, the second type of hybrid approach, S^pM, is more likely to get similar results to SCA2 as q increases because the higher q, the more are iterations of this type of hybrid approach like SCA2 iterations. So MS gets the best balance between accuracy and simplicity for simple uniform structure.

11.2.3 Simple intermediate structure

In Section 4.5.3, it was shown that for the simple intermediate structure, no SCA method retrieved the structure from the population variance covariance matrix V_0 when $k = 0$. So, hopefully, hybrid approaches can improve on this.

Hybrid method M^2S gets the same results as those of SCA2 and SCA5 for the first two components (Figure 11.3). The results for the third and fifth components obtained by hybrid

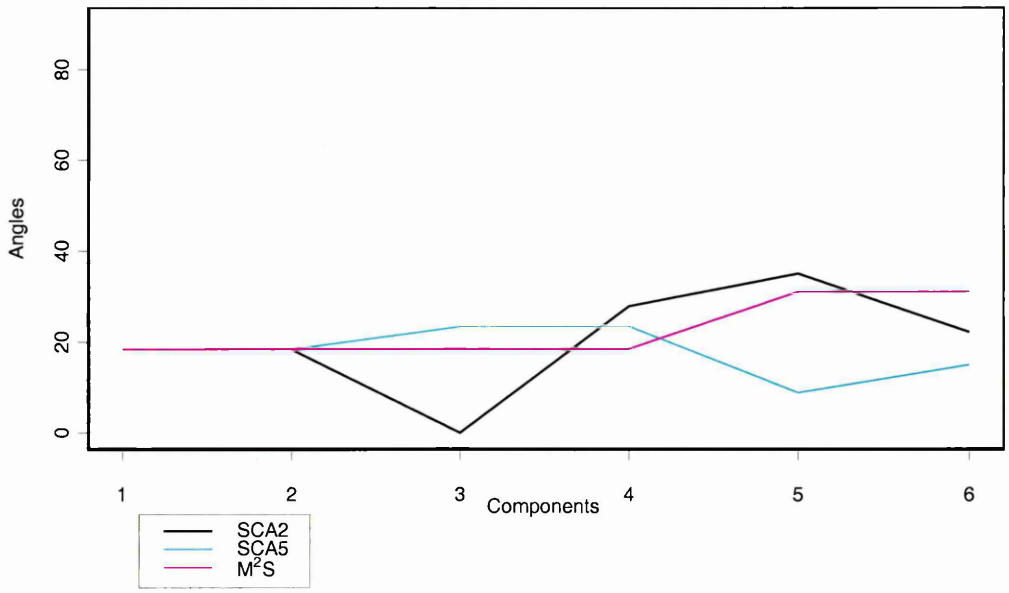

Figure 11.3: The angles obtained by M^2S, SCA2 and SCA5 for 6 dimensional simple intermediate structure

method M^2S are between those of SCA2 and SCA5. The result for the fourth component obtained by hybrid method M^2S is more accurate than both those of SCA2 and SCA5. The result for the last component obtained by hybrid method M^2S is less accurate than those of SCA2 and SCA5. In other words hybrid method M^2S makes some components at least as accurate as the best results of SCA2 and SCA5 but other components less accurate than the best components obtained by SCA2 and SCA5.

All the components produced by SCA2 and hybrid method M^2S (Figure 11.4) are simple. The first five components produced by SCA5 are simple. As in Section 11.2.2, the means of components produced by hybrid method M^2S are just between those of SCA2 and SCA5.

The results obtained by MS and M^3S are less accurate than those of SCA2 for all components and except for component 3 less accurate than those of SCA5. From iteration 4 onwards,

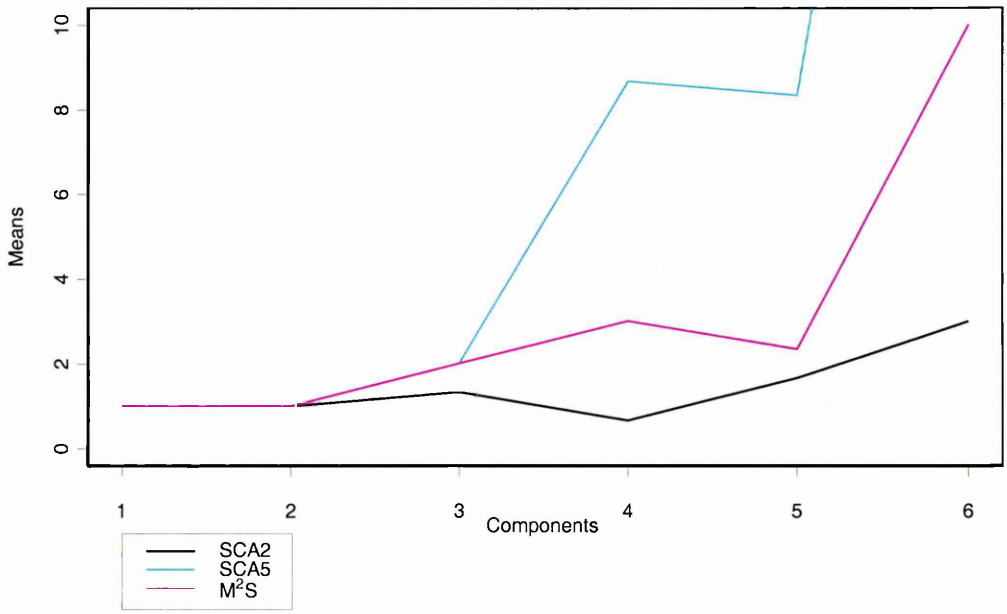

Figure 11.4: The means of the components produced by M^2S, SCA2 and SCA5 for 6 dimensional simple intermediate structure

SCA5 only performs single updates. All the other hybrid approaches, M^qS ($q \geq 4$), get the same results as those of SCA5. So, as q increases, the first type of hybrid approach, M^qS ($q \geq 1$) is more likely to get similar results to SCA5 because more iterations of the first type of hybrid approach M^qS ($q \geq 1$) that like SCA5.

S^2M makes the results for some components more accurate than those obtained by SCA2 and SCA5 and other components less accurate than the best results obtained by SCA2 and SCA5. All the other second type of hybrid approaches S^qM ($q = 1$ and $q \geq 3$) do not improve on the results of SCA2 and SCA5, but they get the same results as SCA2 and SCA5 for the first two components. As q increases, the second type of hybrid approach, S^qM, is more likely to get similar results to SCA2 because more iterations of the second type hybrid approach, S^qM, is equivalent to iterations of SCA2.

All in all, hybrid approaches M^2S and S^2M make the results for components 1, 2 and 4 not less accurate and the results for other components less accurate than the best results obtained by SCA2 and SCA5. The results obtained by M^2S are not less accurate than those obtained by S^2M except for component 4. So it is possible for hybrid approaches to improve the results for simple intermediate structure.

11.2.4 Complex block structure

In this subsection, the application of hybrid approaches to the 6 dimensional data with complex block structure will be discussed. No hybrid approach can improve on the results of SCA2 because SCA2 retrieved the corresponding simple block structure.

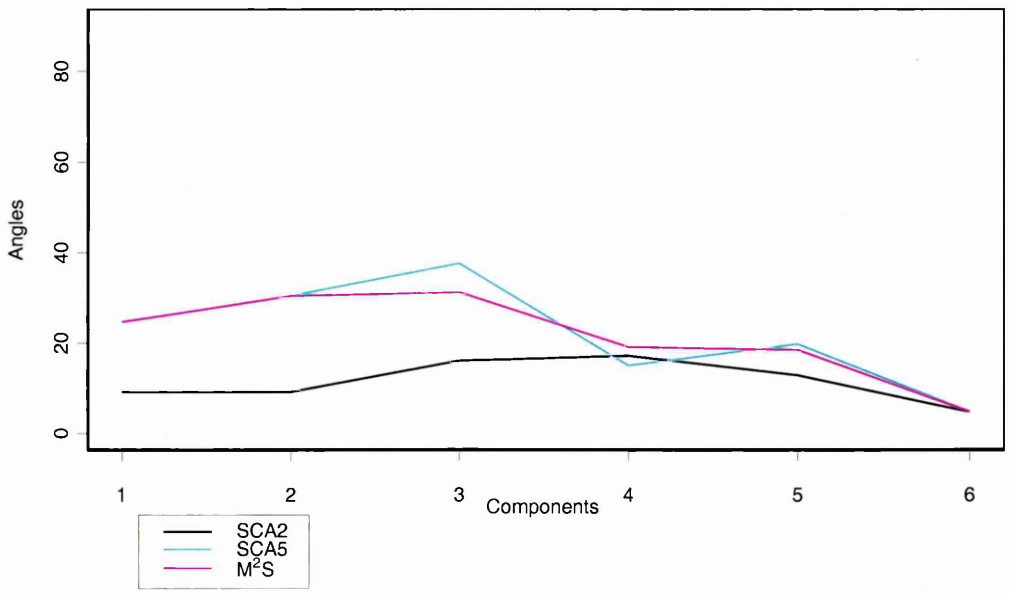

Figure 11.5: The angles obtained by M^2S, SCA2 and SCA5 for 6 dimensional complex block structure

The results obtained by hybrid method M^2S (Figure 11.5) are less accurate than those

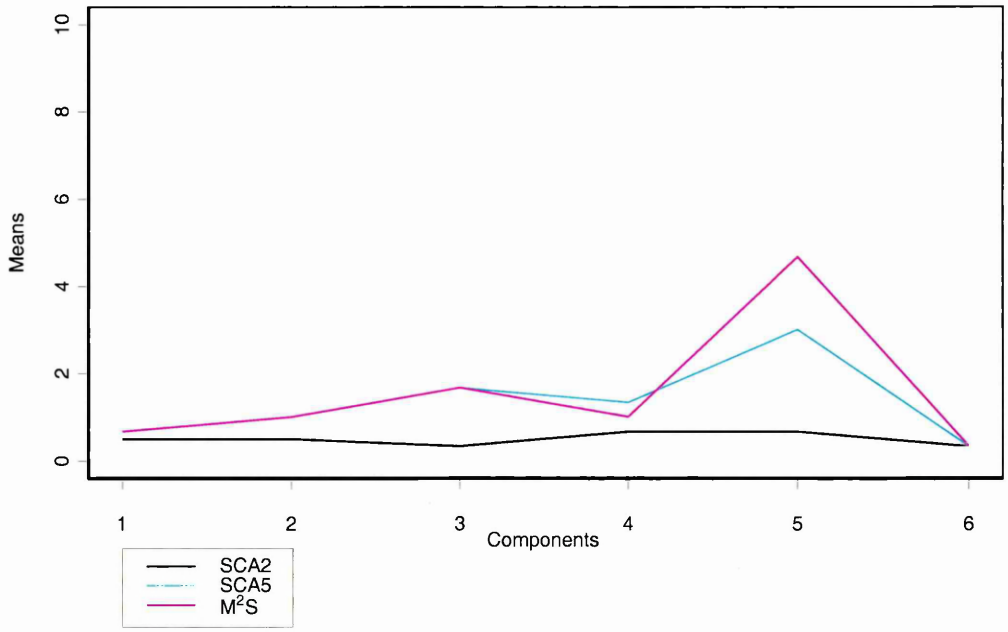

Figure 11.6: The means of the components produced by M^2S, SCA2 and SCA5 for 6 dimensional complex block structure

obtained by SCA2 but, with the exception of component 4 at least as accurate as those obtained using SCA5. All the components produced by hybrid method M^2S, SCA2 and SCA5 are simple (Figure 11.6). The results obtained by SCA2 are more simple than those obtained by SCA5 (Section 4.6.1) and hybrid method M^2S.

Checking the iterations of SCA5, it was shown that only the first and fourth iterations of SCA5 used multiple updating. So the hybrid approaches M^qS, $q = 1, 2, 3$, gave the same results to each other. Hybrid approaches M^qS with $q \geq 4$ obtain the same results as those of SCA5 because SCA5 only took 4 iterations. So the results for the first type of hybrid approach M^qS, are the same or very similar to those of SCA5.

SCA2 took 5 iterations. So only hybrid approaches S^qM with $q = 1, \ldots, 4$ are worth considering. When $q \geq 5$, the second type of hybrid approach S^qM is the same as SCA2. The hybrid

approach SM gave the same results as SCA2. But this hybrid approach took 4 iterations, only the second of which was essentially a multiple updating step. S^qM with $q \geq 2$ led to the results being the same as that of SCA2 and took the same number of iterations. In other words, the second type of hybrid approach S^qM with $q \geq 1$ get the same results as SCA2.

So, as expected, as q increases, the results obtained by the first type of hybrid approach, M^qS is the same or very similar to those obtained by SCA5 because more iterations are like SCA5 iterations, and the results obtained by the second type of hybrid approach, S^qM is the same as SCA2.

11.2.5 Complex uniform and intermediate structures

In this subsection, the application of hybrid approaches M^qS and S^qM with $q \geq 1$ to 6 dimensional complex uniform and intermediate structures will be investigated.

For complex uniform structure, it is possible for hybrid approaches to improve the results obtained by SCA2 and SCA5 because SCA2 and SCA5 did not retrieve the corresponding simple uniform structure (Section 4.6.2) and generally the simple structure is the closest simple approximation to corresponding complex structure by construction. So there is room for hybrid approaches to get more accurate results than those of SCA2 and SCA5.

For 6 dimensional complex uniform structure, the results obtained by hybrid method M^2S (Figure 11.7) are more accurate than those obtained by SCA2 for the first three components and more accurate than those obtained by SCA5 for all the components. The results obtained by SCA2 are more accurate than those obtained by M^2S and SCA5 for the last three components. All the components produced by hybrid method M^2S and SCA5 (Figure 11.8) are simple. Only the first two components produced by SCA2 are simple, and other components produced by SCA2 are complex. So the results obtained by hybrid method M^2S are better than those obtained by SCA2 and SCA5. Except for components 3 and 6 the results obtained by hybrid approach MS are the same as those obtained by hybrid method M^2S. Hybrid approaches M^qS with $q \geq 3$ get the same results as the hybrid method M^2S.

The second type of hybrid approach S^qM was also investigated. I only consider S^qM when $q = 1, \ldots, 8$ because SCA2 took 9 iterations. The results obtained by the second type of hybrid

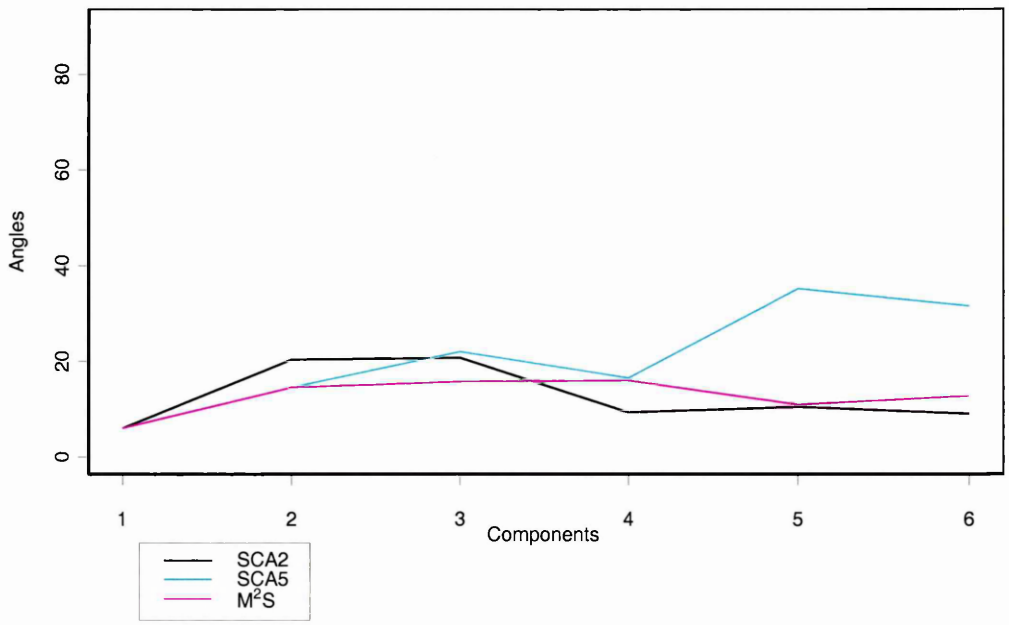

Figure 11.7: The angles obtained by M^2S, SCA2 and SCA5 for 6 dimensional complex uniform structure

approach SM are much more accurate than those obtained by SCA2 and SCA5 except for component 6. Even for component 6, the results obtained by hybrid approach SM are only slightly less accurate than those obtained by SCA2 but more accurate than those obtained by SCA5. Actually, SM just retrieves the simple uniform structure. So the results obtained by SM are better than those obtained by M^2S, SCA2 and SCA5 except for component 6. This is different from what has been found for other structures. As said before in this section it is possible for hybrid approaches to improve the results obtained by SCA2 and SCA5. Indeed the first two components are the same as the first two components of the simple uniform structure, and it only took 4 iterations. The results of other hybrid approaches S^qM with $q > 1$ are no more accurate than those of hybrid approach SM.

For the complex intermediate structure, hybrid method M^2S, SCA2 and SCA5 get the same

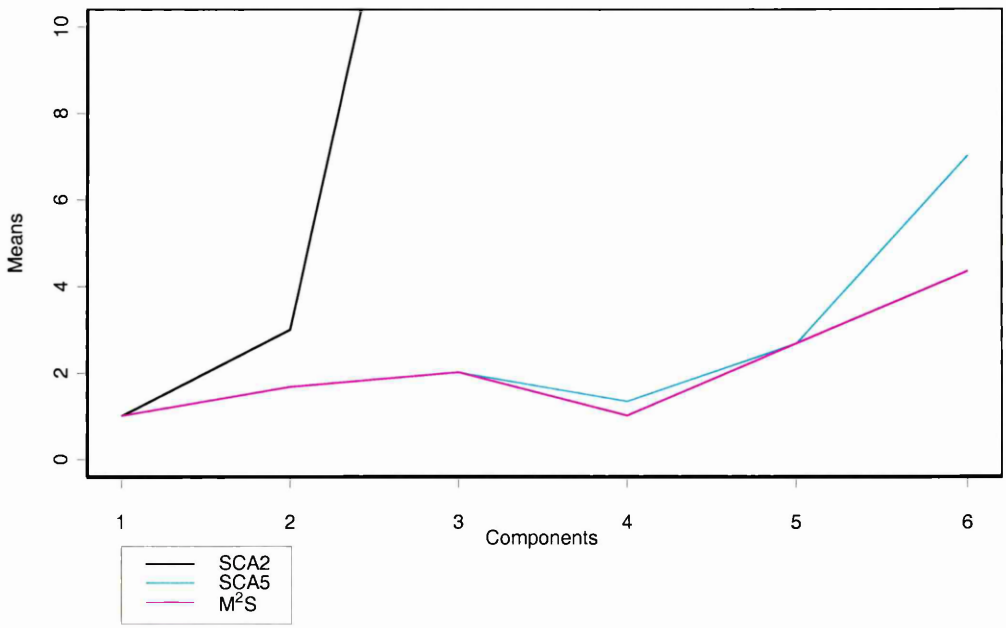

Figure 11.8: The means of the components produced by M^2S, SCA2 and SCA5 for 6 dimensional complex uniform structure

results for the first two components. The results obtained by M^2S are more accurate than those of SCA2 and SCA5 for some components and less accurate than those of SCA2 and SCA5 for other components.

To sum up the application of hybrid approaches M^qS and S^qM ($q \geq 1$) to 6 dimensional data, there is at least one hybrid approach (M^qS and S^qM with $q = 1$ and 2) whose results are the same or better than the best results of SCA2 and SCA5. So overall hybrid approaches M^qS and S^qM with $q = 1$ and 2 are slightly better than SCA methods. In the next section, I will check whether hybrid approaches can improve on the results obtained by SCA methods for 8 dimensional data.

11.3 Hybrid approaches for 8 dimensional data, $k = 0$

In this section, the application of hybrid approaches to 8 dimensional data will be discussed. As in Section 11.2, two types of hybrid approaches M^qS and S^qM $(q \geq 1)$ will be implemented. Also as in the previous section, I just consider the population results. The eigenvectors matrices for the 8 dimensional data used in this section were given in Section 6.2.

11.3.1 Simple structures

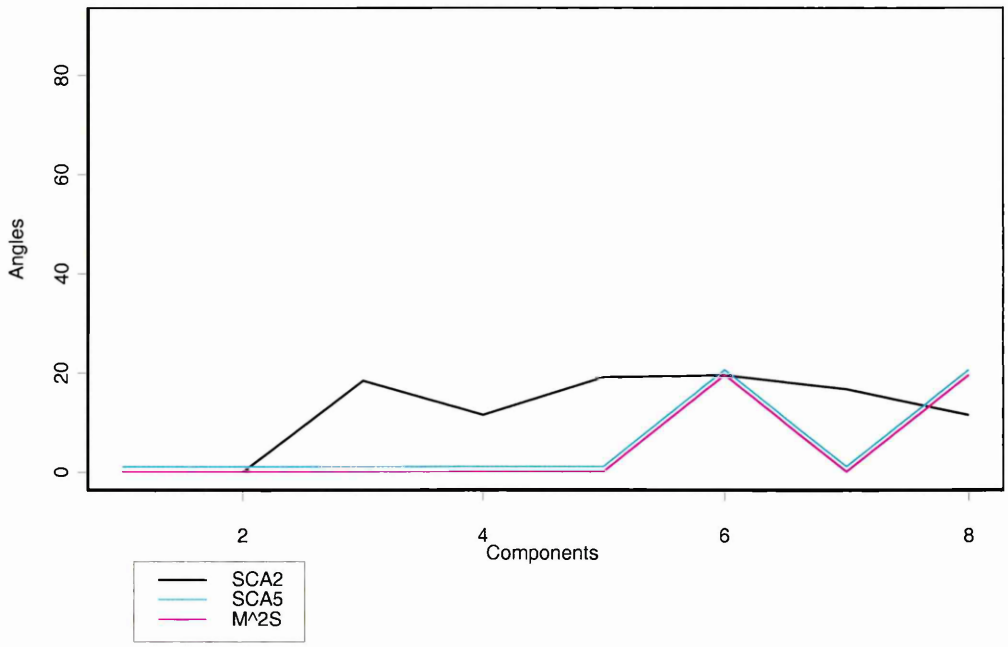

Figure 11.9: The angles obtained by M^2S, SCA2 and SCA5 for 8 dimensional simple block structure

First, consider the simple block structure. Except for the last component the results obtained by the hybrid method M^2S are more accurate than those obtained by SCA2 (Figure 11.9). The results obtained by hybrid method M^2S are the same as those obtained by SCA5, to distinguish

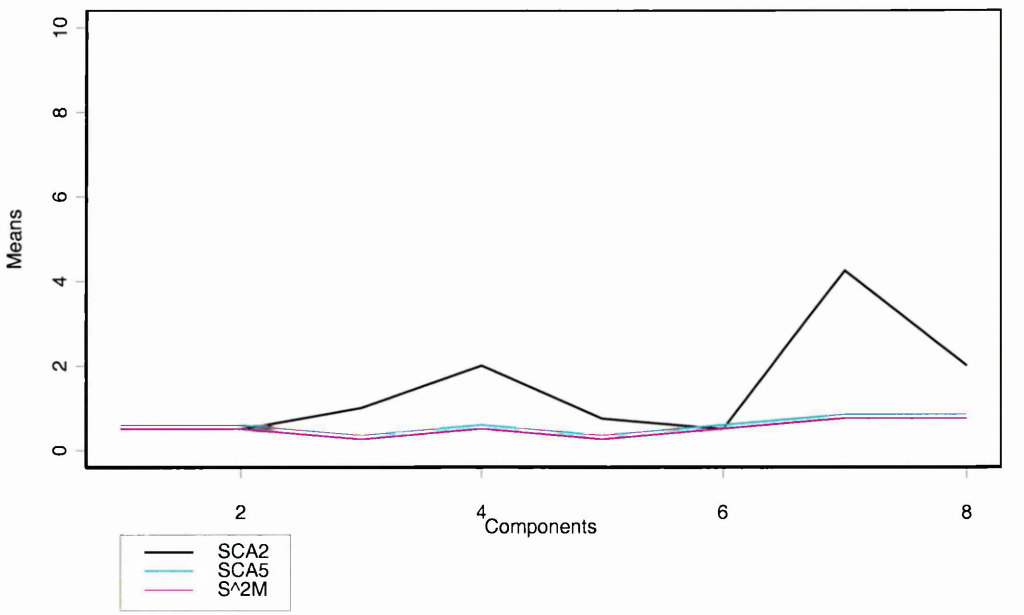

Figure 11.10: The means of the components produced by M^2S, SCA2 and SCA5 for 8 dimensional simple block structure

the difference between SCA5 and M^2S, I add 2° for each angle for SCA5. It is very clear in Figure 11.9 that only the results obtained by SCA5 for components 6 and 8 could possibly be improved by hybrid approaches. All the components from M^2S, SCA2 and SCA5 are simple (Figure 11.10). The components obtained by SCA5 (add 0.1 for the mean of each component for SCA5) and hybrid approach M^2S are more simple than those obtained by SCA2. This is a special case. Generally the components obtained by SCA2 are more simple than those produced by SCA5 and hybrid approach M^2S.

The reason M^2S and SCA5 gave same results is because for the first 2 iterations, SCA5 and hybrid method M^2S are going to be the same by definition. But additionally iterations 3 and 4 of the hybrid approach turned out to be equivalent to iteration 3 of SCA5 and iteration 5 of hybrid approach M^2S turned out to be the same as iteration 4 of SCA5.

Now, consider the first type of hybrid approach M^qS with $q \geq 1$. As expected, the hybrid approaches M^qS with $q \geq 3$ get the same results as those obtained by SCA5 because SCA5 took 4 iterations and iteration 4 was effectively single updating. However, the results obtained by hybrid approach MS are also the same as those obtained by SCA5. So hybrid method M^qS with $q \geq 1$ always ended up the same as SCA5 for the simple block structure.

The results obtained by the second type of hybrid approach S^qM with $q \geq 5$ are the same as those obtained by SCA2. The results obtained by S^qM, $q = 1, \ldots, 4$ are the same each other and more accurate than those obtained by SCA2. The results obtained by S^qM, $q = 1, \ldots, 4$, are not less accurate than the best results obtained by SCA2 and SCA5 except for components 5 and 8.

For the 8 dimensional simple uniform and simple intermediate structures, SCA2 and SCA5 got the same results, and all the hybrid approaches M^qS and S^qM with $q \geq 1$ gave the same results as SCA2 and SCA5.

11.3.2 Complex structures

Consider the application of hybrid approaches to complex block structure. Except for component 6, the results obtained by hybrid method M^2S and SCA5 (add 2° for each angle)are at least as accurate as those of SCA2 (Figure 11.11). The results obtained by hybrid method M^2S are the same as those of SCA5. All the components hybrid method M^2S, SCA2 and SCA5 (add 0.1 for each mean) are simple (Figure 11.12).

The reason why M^2S and SCA5 produce the same results is as follows. Remember that SCA5 took 4 iterations and SCA2 took 10 iterations. M^2S took 5 iterations. The first three iterations of SCA5 are multiple updating and the last iteration of SCA5 is single updating. The first two iterations of hybrid method M^2S and SCA5 are the same. This follows from the definition of hybrid approach. The iterations 3 and 4 of the hybrid method M^2S are the same as iteration 3 of the SCA5. So, the last iteration of hybrid method M^2S is the same as the last iteration of SCA5.

As expected, the first type of hybrid approach M^qS with $q \geq 3$ obtain the same results as SCA5 and take the same number of iterations.

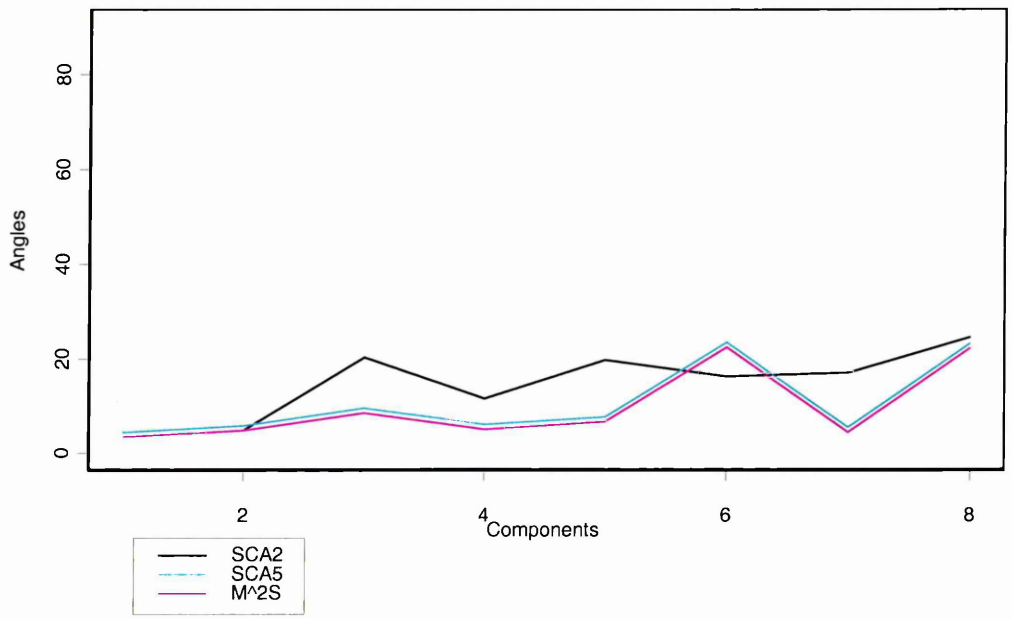

Figure 11.11: The angles obtained by M^2S, SCA2 and SCA5 for 8 dimensional complex block structure

The results obtained by all the second type of hybrid approach S^qM ($q \geq 1$) are less accurate than those of SCA5. This is not surprising because this type of hybrid approach has similar results to those of SCA2, and the results obtained by SCA2 are less accurate than those obtained by SCA5. The results obtained by hybrid approaches S^qM with $q \geq 4$ are the same as those obtained by SCA2. So this type of hybrid approach is generally less accurate than the first type of hybrid approach M^qS ($q \geq 1$) except for component 6.

For the 8 dimensional complex uniform structure, the results obtained by hybrid method M^2S are the same as those obtained by SCA5 and SCA2. But this hybrid method needs more iterations than SCA5. This is because by definition this hybrid approach only has two multiple updating iterations. The hybrid approach M^qS with $q \geq 1$ and S^qM with $q \geq 1$ get the same results as SCA2 and SCA5.

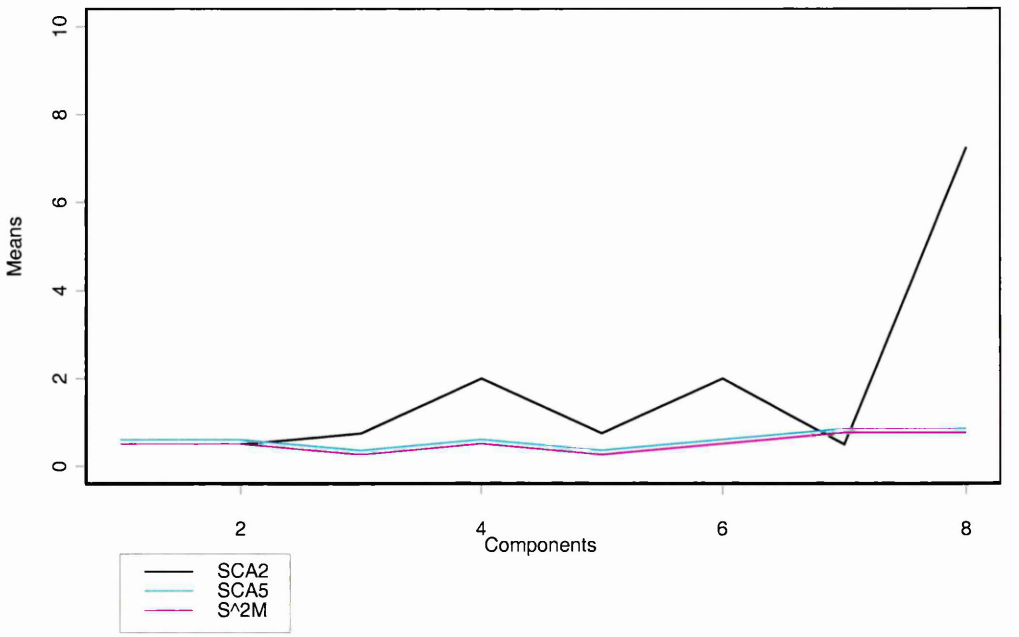

Figure 11.12: The means of the components produced by M^2S, SCA2 and SCA5 for 8 dimensional complex block structure

Finally, consider the 8 dimensional complex intermediate structure. The results obtained by hybrid approach M^2S are the same as those of SCA5 and SCA2. In fact, the results obtained by all the hybrid approaches are the same as those obtained by SCA2 and SCA5. This is expected because for the complex 8 dimensional intermediate structure the first type of hybrid approach M^qS $(q \geq 1)$ gets the same results as those obtained by SCA5, the second type of hybrid approach S^qM $(q \geq 1)$ gets the same results as those obtained by SCA2, and SCA2 and SCA5 gave the same results.

So for 8 dimensional data, only the results obtained by SCA methods for 8 dimensional simple block structure are improved by hybrid approach S^qM when $q \geq 2$ but by not very much. This is because SCA methods did very well for 8 dimensional data. Even so, the results obtained by hybrid approach M^2S are not worse than the best results obtained by SCA2 and SCA5. So

hybrid approaches are very good for 8 dimensional data. Recall that SCA methods also did very well for 10 dimensional data (Section 6.4), it is expected that the results obtained by SCA methods for 10 dimensional data will not be improved by very much by hybrid approaches. So I will not discuss the population results obtained by hybrid approaches for 10 dimensional data in this chapter.

11.4 Hybrid approaches for data in practice, $k = 0$

In practice, the actual eigenvectors structures generally are a mixture of complex block, uniform and intermediate structures. So hybrid approaches should be useful for improving the results obtained by SCA methods or get the best balance of accuracy and simplicity. To explore this, I will apply the hybrid approaches to the data introduced in Section 1.3. It is known that when $k = 0$, the results obtained by SCA2 or SCA5 were very good and gave the best balance between approximation and interpretation for the RI data and the sparrow data (Sections 3.5.1 and 3.5.2). So hybrid approaches M^qS and S^qM with $q = 1$ and 2 are only applied to the employment in Europe data (Section 1.3.3) and the Jeffers' pitprop data (Section 1.3.4).

11.4.1 Employment in Europe data

As in Section 1.3.3, I only consider here the first three components of employment in Europe data. The results obtained by SM, MS and M^2S for this data are less accurate than those obtained by S^2M. So I just give the graphs of S^2M, SCA2 and SCA5.

The result obtained by S^2M (Figure 11.13) is less accurate than that obtained by SCA2 and very similar to that obtained by SCA5 for the first component, the results obtained by S^2M are much more accurate than those obtained by SCA2 and SCA5 for components 2 and 3.

The components produced by S^2M, SCA2 and SCA5 are simple for the first three components (Figure 11.14). So the results of S^2M are much better than those obtained by SCA2 and SCA5 for components 2 and 3.

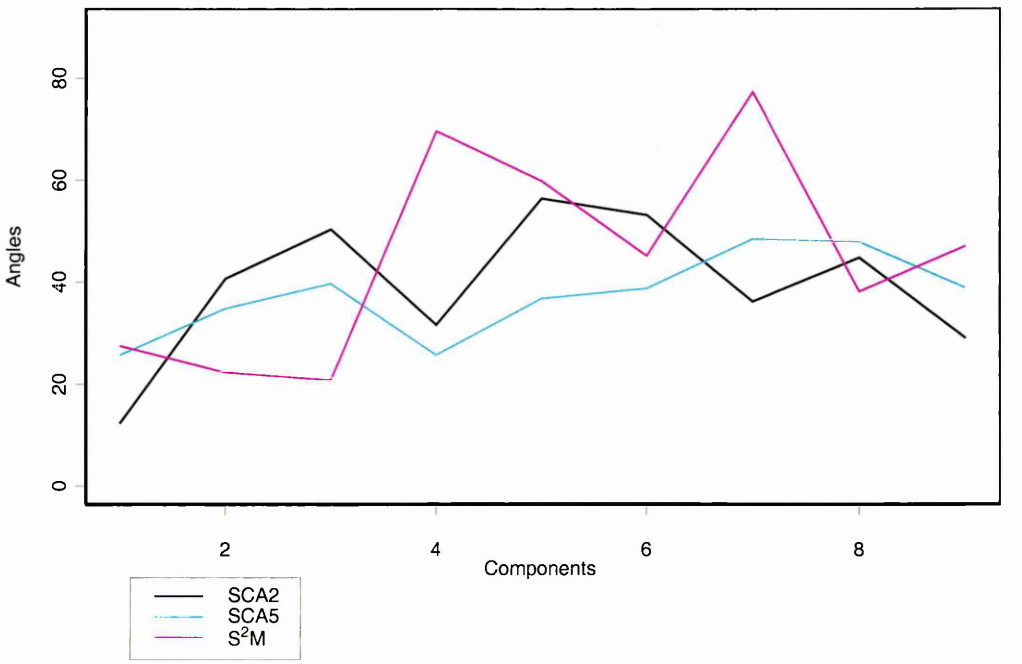

Figure 11.13: The angles obtained by S^2M, SCA2 and SCA5 for the employment in Europe data

11.4.2 Jeffers' pitprop data

As in Section 1.3.4, I just consider here the first four components of the Jeffers' pitprop data. First consider the population results obtained by SM applied to the Jeffers' pitprop data.

The result obtained by SM (Figure 11.15) is more accurate than the results of both SCA2 and SCA5 for component 2, the results obtained by SM are between those obtained by SCA2 and SCA5 for components 1 and 3 and very similar to the result obtained by SCA2 and more accurate than that obtained by SCA5 for component 4.

Only the first two components produced by SM are simple (Figure 11.16). The components produced by SCA2 are the simplest for all the components. So SM improves only on the results of SCA2 and SCA5 for component 2.

The results obtained by S^2M, MS and M^2S are not better than those obtained by SM.

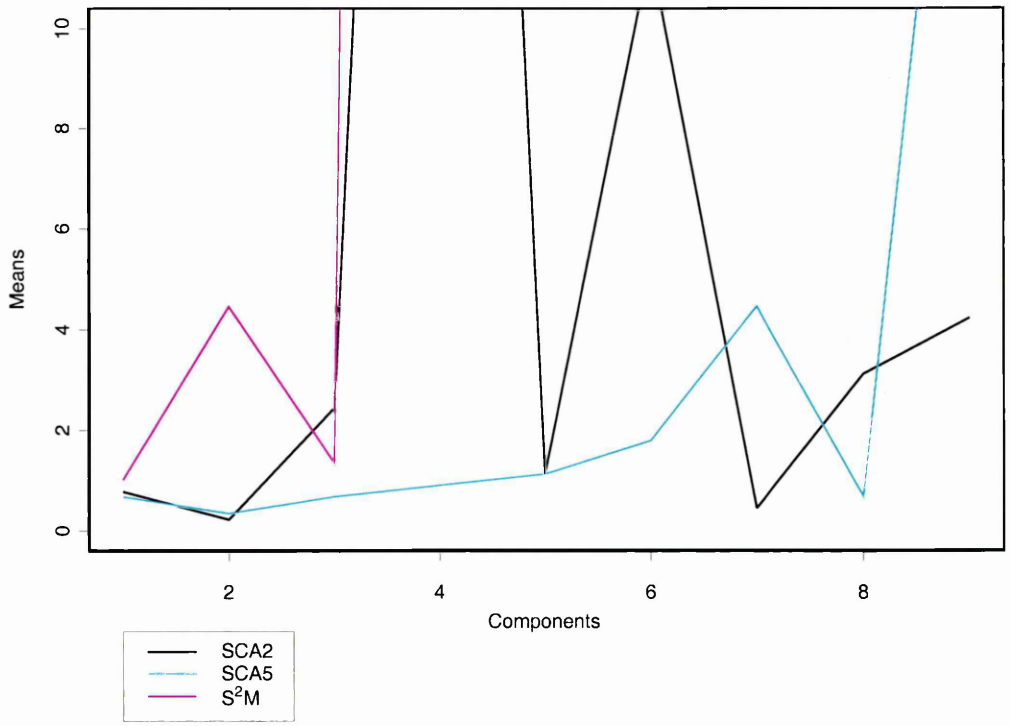

Figure 11.14: The means of the components produced by S^2M, SCA2 and SCA5 for the employment in Europe data

So the details will not be given for these three hybrid approaches here. Hybrid approaches can improve on the results obtained by SCA2 and SCA5 for some components of this data.

11.5 Discussion and conclusion

Hybrid approaches use different SCA methods with possibly different restrictions and different values of k in different iterations. Two types of hybrid approaches, both using the criterion maximal improvement in variance and $k = 0$ at all iterations were considered. One type of hybrid approach M^qS with $q \geq 1$ used q multiple updating iterations followed by single updating for all remaining iterations. The other type of hybrid approach S^qM with $q \geq 1$ used single updating for the first q iterations followed by multiple updating for the remaining iterations. As

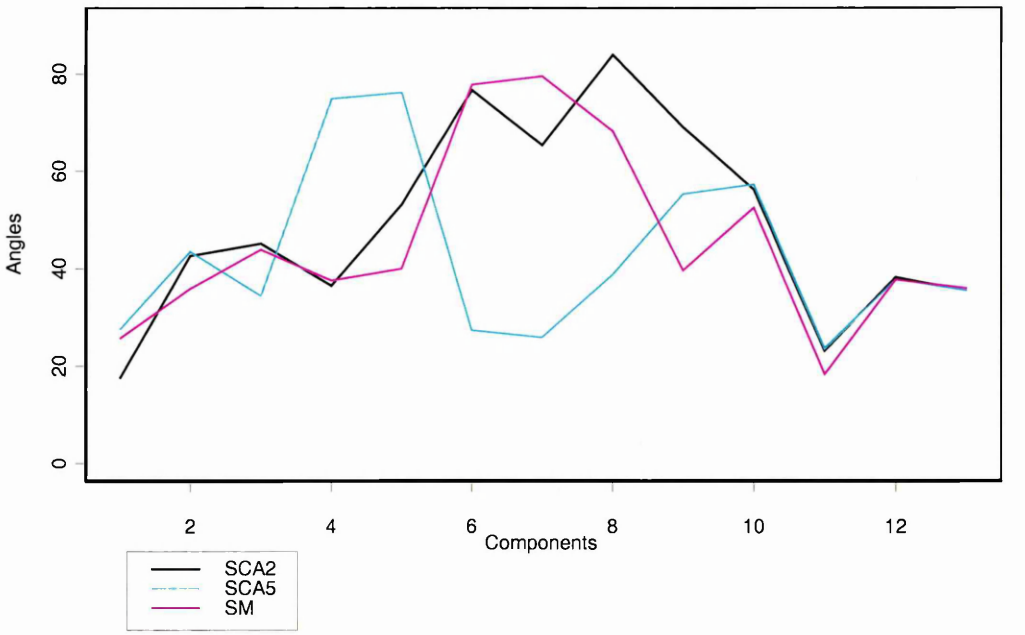

Figure 11.15: The angles obtained by SM, SCA2 and SCA5 for the Jeffers' pitprop data

q increases, the first type of hybrid approach, M^qS, is more likely to get similar results to SCA5, whereas the second type of hybrid approach, S^qM is more likely to get similar results to SCA2. This is because the higher the q, the more iterations of M^qS are like SCA5 iterations, and the more iterations of S^qM are like SCA2 iterations. So generally it is enough just to consider the hybrid approaches M^qS and S^qM with $q = 1$ and 2.

In general, the hybrid approaches M^qS and S^qM with $q = 1$ and 2 are at least as accurate as those obtained by SCA2 and SCA5. For example, for the 6 dimensional complex uniform structure, the results obtained by hybrid method SM gets the corresponding simple uniform structure back exactly and are much more accurate than the results obtained by SCA2 and SCA5 except for component 6. If I have to apply one of SCA2, SCA5 and hybrid method SM to data in practice. I would like to use SCA2 first, if the results are not good enough, the next method I might try is SM.

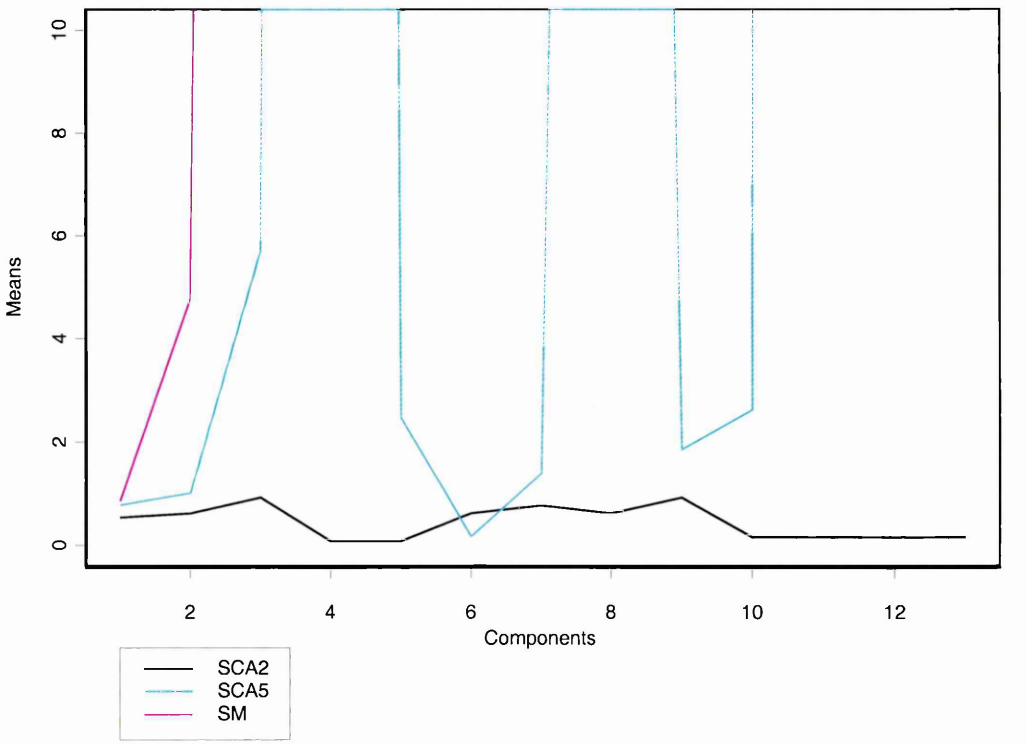

Figure 11.16: The means of the components produced by SM, SCA2 and SCA5 for the Jeffers' pitprop data

Chapter 12

Discussion and further work

The thesis has mainly focused on the simple components algorithm (Vines, 2000). As in principal components analysis (PCA), the loading vectors of simple components (SCs) are orthogonal with each other. However the simple components algorithm (SCA) achieves components that generally are easier to interpret. In this chapter I am going to summarize the conclusions given in this thesis. Section 12.1 will summarize principal component analysis and approaches used to modify PCs to enhance the interpretation of PCs. Section 12.2 gives a summary of the simple components algorithm. Section 12.3 gives a summary of population results and sample simulation results for various SCA methods. Section 12.4 gives the conclusions of the population results using different values of k and different sets of eigenvalues. Section 12.5 sums up the population results using different restrictions. Section 12.6 discusses the conclusions for combined results and hybrid approaches. Finally, Section 12.7 gives overall conclusions and suggestions for further work.

12.1 Principal components and the approaches to modify principal components

Chapter 1 dealt with principal component analysis (PCA). Some of the examples showed that the exact loadings of principal components can make interpretation of PCs difficult, even if

the variables with small loadings were ignored and loadings were rounded. So various kinds of approaches have been pursued for the simplification and interpretation of principal components.

Chapter 2 reviews approaches that have been pursued for the simplification and interpretation of principal components. Rotation is a widely used technique in practice because it is readily available in software packages and it can provide useful simplification of PCs in some cases. Another common approach is truncation of loadings, that is to take any loadings less than some threshold value as zero. Truncation of loadings can be misleading (Cadima and Jolliffe, 1995). The variables with small loadings are not necessarily unimportant. Other approaches such as SCoTLASS, SCA_{RG} and SPCA etc are not popularly adopted in practice because these methods are not easy to use and are much newer so might not have had enough time to catch on. The interpretation of modified PCs produced by these three techniques is generally much easier than that of corresponding PCs and the variance explained by modified PCs is usually slightly less than that of PCs. However ignoring the size of angles between the modified components and principal components generally makes the patterns of the modified components obtained by SCoTLASS, SCA_{RG} and SPCA much different to those of the corresponding principal components.

The aims of all the techniques discussed in Chapter 2 are to get components that are easy to interpret according to a definition of simplicity. Most of these techniques do not use an explicit definition of simplicity of the loadings, and do not care about the size of the angles between the modified PCs and corresponding PCs. As was said in Section 1.2, PCs have two nice properties, they are orthogonal and uncorrelated. To enhance interpretation of PCs, the techniques which modify PCs have to sacrifice at least one of these two properties. I prefer the techniques which have explicit definition of simplicity and keep one of the two properties of PCA. Some techniques, such as the technique by Chipman and Gu's technique (2003), have a clear definition of simplicity and consider the size of angles between the modified PCs and original PCs. Other techniques, such as Rousson and Gasser's technique (2004), have a distinct definition of simplicity but do not consider the size of angles. Unfortunately, these two techniques lose both of the two nice properties of PCs. This makes the computation of the amount of variance of each component accounts for difficult (Section 1.2).

12.2 Simple principal components

Chapter 3 described the simple principal components algorithm (SCA). SCA (Vines, 2000) works from the variance covariance or correlation matrix directly. The algorithm consists of a series of simple linear transformations. Each transformation searches a restricted set of directions within a two-dimensional subspace for the directions with which the data have maximal variance or maximal improvement in variance. SCA produces integer valued loadings vectors. Simple components are orthogonal with each other. The integer valued loadings vectors approximate the loadings vectors obtained via PCA, but are hopefully more interpretable. The Jacobi method was also introduced in Chapter 3. The Jacobi method is used to transform a real symmetric matrix arbitrarily close to a diagonal matrix by a sequence of orthogonal transformations. In Chapter 3, I also pinned down the link between the SCA algorithm and the Jacobi method, and hence explained why the SCA can be called a Jacobi-like method. I also got some useful results about the difference (α) between the SCA angle and the Jacobi angle (θ), and quantified the relationship between angles α and θ. It is proved in Section 3.7 that for any given directions r and s, the SCA angle at each step is the one with minimal $|\alpha|$. So in every step, the SCA angle is the one with minimal $|\alpha|$. However what has not been shown is that out of all pairs of directions to transform, the best pair to choose is the one for which $|\alpha|$ is minimized.

The results from SCA were compared with the principal components by applying it to the same data sets as in Chapter 1. It was shown that simple components can be very good approximations to the PCs for some data sets. For example, for the RI data, the angles between simple components and their corresponding principal components are very small. Also each simple component accounts for a similar amount of variance compared to the corresponding PC. Furthermore, the simple components produced are in general easier to interpret than the equivalent PCs because of the small valued integer loadings.

The performance of SCA is controlled by a nonnegative integer k, which determines the number of directions considered for each simplicity preserving transformation at each iteration. As k increases, the directions available to get simple components increases. For the examples in Chapter 3 $k = 0$ is a good choice.

Six different variations of the SCA algorithm such as multiple updating with maximal variance or maximal improvement in variance or single updating with maximal variance or maximal improvement in variance were introduced in this chapter. From the results obtained from different SCA methods, it was found that different simple components can be obtained for the same data set. So these six different SCA methods were investigated in the following chapters to check which SCA method is good.

12.3 Population results and sample simulation results

In Chapters 4 and 6, the performances of the different SCA methods when $k = 0$ were investigated by comparing the results of SCA methods applied to a variance covariance matrix V_0. V_0 has exact eigenvectors of known structure in Sections 4.3 and 6.2. V_0 is variance covariance matrix after the columns of eigenvectors structures in Sections 4.3 and 6.2 are normalized, the corresponding eigenvalues was also given in Sections 4.3 and 6.2. The results applied to V_0 are called population results. The population results of PCA must be the original eigenvectors structures in Sections 4.3 and 6.2, so the population results of PCA were not considered here.

In Chapters 5 and 7, both the performances of the different SCA methods when $k = 0$ and PCA were investigated by comparing sample simulation results. Here a sample simulation is based on 500 data sets, each data set having 500 observations. Each data set was assumed to have a normal distribution $N(0, V_0)$.

The population results and the sample simulation results generally gave the same conclusion. This is a useful feature because it is easier to investigate the population results than the sample simulation results. So, generally, in order to know which SCA method is more accurate or more simple, it is enough just to consider the population results. In general, SCA2 (single SCA method with maximal improvement in variance), SCA5 (multiple SCA method with maximal improvement in variance) and SCA6 (multiple SCA method with maximal variance) gave the best results whatever the structures of data. Generally the results obtained by SCA methods

with maximal improvement in variance were at least as accurate but less simple than those obtained by SCA methods with maximal variance. The results obtained by the single SCA method with maximal improvement in variance (SCA2) were at least as accurate and more simple than those obtained by the multiple SCA method with maximal improvement in variance. However, the results obtained by the multiple SCA method with maximal variance were more accurate but less simple than the results obtained by single SCA methods with maximal variance. The condition $l_r = l_s$ made the results obtained by the SCA methods no more accurate but more simple than the results obtained by SCA methods without this condition. The explanation of above conclusions had been given in Section 4.7.

The results obtained by SCA methods generally are very good for block structure and uniform structure. In particular SCA2 recovered simple block and simple uniform structures. The simple components algorithm was not so good for intermediate structure.

If the number of observations in a data set was increased, it was shown that the sample simulation results obtained by SCA methods improved greatly. This is expected because as the size of the sample increases, the estimated parameters are more accurate, i.e. the elements of the variance covariance matrix of the sample are closer to the elements of V_0. So the variance covariance matrix of the sample is closer to the variance covariance matrix V_0 when the sample size is increased.

Generally, whatever the dimension of the data, for simple block and simple uniform structures, the most accurate sample simulation results obtained by SCA were more accurate than the sample simulation results obtained by PCA.

For large dimensional data, generally for complex block and uniform structures, the most accurate sample simulation results obtained by the SCA method were more accurate than the sample simulation results obtained by PCA. The reason why this is true for simple block and simple uniform structures was given above. For large dimensional complex block and complex uniform structures, it is true because the angles between the simple structure and corresponding complex structure are smaller than the difference between the sample simulation results of PCA and SCA for simple block and simple uniform structures.

For 6 dimensional complex structures and any dimensional intermediate structures, the sam-

ple simulation results were generally slightly less accurate than those obtained by PCA but were more simple to interpret. However, taking into account the generally easier interpretation of the simple components, the angles between simple components and corresponding principal components are very small for complex block and complex uniform structures. So if the results obtained by PCA are good for block and uniform structures, generally the results obtained by SCA should be good for block and uniform structures too.

SCA methods are not very good for intermediate structure because of the bigger angles for all the examples, one reason is that $k = 0$ is not a good choice for this structure. So only the population results of SCA methods when $k = 0$, 1 and 2 were investigated in Chapter 8 because generally the population results and the sample simulation results gave the same conclusions.

12.4 Further population results

In Chapter 8, the population results of SCA2 using different k (k = 0, 1 and 2) and using three different sets of eigenvalues were investigated. It was shown in Section 3.3.2 that as k increases, there are more directions for SCA methods to consider for each simplicity preserving transformation at each iteration. Also the directions given by $k + 1$ include all the directions given by k. So more accurate results are expected for all the SCA methods as k increases. In fact, the results obtained by SCA2 with larger k are not necessarily more accurate than those with smaller k. This is because generally the maximal improvement in variance just makes the results the most accurate at each step. The maximal improvement in variance of SCA2 with larger k at each step is not always more than the maximal improvement in variance with smaller k because from some step the variance covariance matrix is different for different k. Even if the maximal improvement in variance of SCA2 with large k is more than that of SCA2 with smaller k at each step, this can not guarantee the most accurate results finally (an example of this was given in Section 8.2).

Furthermore the results with larger k are not as simple as those with smaller k. This is because the larger the k, the longer the new directions in general (Sections 3.3.2, 3.3.3 and 8.2.1). Also as k increases, generally the steps taken by SCA2 increase and more steps also

mean that the components are longer generally. So the results of SCA methods with larger k are usually not better than those with smaller k because they are not as simple as those with smaller k.

It was known in Chapters 4 to 7 that when $k = 0$, SCA was not good for intermediate structure. However, when $k = 1$, it was shown that SCA2 generally retrieved the first two components of simple intermediate structure. Also the angles obtained by SCA2 for the complex intermediate structure were just the angles between the first two components of simple intermediate structure and complex intermediate structure. These conclusions are the same as those for block and uniform structures when $k = 0$. So SCA methods can get all the simple structures back exactly if a suitable k is chosen. For complex structures the best angles obtained by SCA methods generally are obtained when the SCA method gets the simple structure back exactly because the simple structure is the closest simple approximation of the corresponding complex structure.

In Chapters 5 and 7, for simple block and simple uniform structures when $k = 0$ it was concluded that the sample simulation results obtained by SCA are better than those obtained by PCA. The population results in Chapter 8, show that the results obtained by SCA2 are the same as the first two components of simple intermediate structure when $k = 1$. This is the same conclusion as SCA2 when $k = 0$ for simple block and uniform structures. So I conclude that the sample simulation results obtained by SCA methods should be better than the sample simulation results obtained by PCA for simple intermediate structure when $k = 1$ (the reason is similar to that given in Section 5.5). So for any simple structure, the sample simulation results obtained by SCA algorithm should be better than those obtained by PCA if a suitable k is used. That is to say, whatever the dimension of the simple structures the sample simulation results obtained by SCA methods should not only be easier to interpret but also more accurate than those obtained by PCA. When an individual SCA method gets the simple intermediate structure for both simple and complex intermediate structures, and the angles between the simple intermediate structure and complex intermediate structure are smaller than the difference between the sample simulation results between PCA and this SCA method for simple intermediate structure when $k = 1$, the sample simulation results of this SCA method for

the corresponding complex intermediate structures are better than those of PCA when $k = 1$. This means that whatever the structure of the data, if PCA is good for some data, SCA is also generally good so long as a suitable k is chosen. In general, the best results appear to be obtained when k equals 0 or 1.

It was also shown in Chapter 8 that generally the eigenvalues of a data set influence the simulation results but not by very much if the same eigenvectors are used. For the simple block structure all the results for the first two components were the same. So it seemed that the eigenvalues of the variance covariance matrix had no effect on the results of simple block structure, though surely this is not true for all possible choices of eigenvalues. For the other structures, the results of SCA methods applied to the variance covariance matrices that had different eigenvalues generally differed but not by very much. In general, when there is a big difference between the first two eigenvalues, the results were more accurate than those obtained when there was an extremely small difference between the first two eigenvalues. This is expected. If the first two eigenvalues have an extremely small difference, the first two principal components have almost equal variance. It is known that the subspace spanned by the eigenvectors which have exactly equal eigenvalues, that are well-separated from all other eigenvectors, is well defined and stable. But these two eigenvectors are unstable and are not uniquely defined (Jolliffe (2002b) Sections 2.4, 3.4, 3.7 and 10.3). So if the first two eigenvalues have an extremely small difference, the population results are sometimes unstable. When there is a big difference between the first two eigenvalues, PCA generally retrieves the corresponding eigenvectors of the population variance covariance matrix, and then for sample simulation data, PCA is very good. The simple components are very close to the corresponding principal components, so the usual properties of principal components apply. This means that two simple components are unstable and not uniquely defined if the first two eigenvalues are exactly equal. The results of SCA should be more accurate if the first two eigenvalues have big difference. However the eigenvector structures are the main factor affecting the results of SCA methods.

Chapter 8 only investigated the effect of k and effect of the sets of eigenvalues on the results of SCA2. Of course the same issues can be investigated for all the other SCA methods, but considering the small difference between the different SCA methods, it is expected that the

conclusions obtained by SCA2 may be extended to other SCA methods. Also, in Chapter 8, only population results were investigated. Of course, sample simulation results could also be considered. But it is expected that these would lead to the same conclusions as for the population results.

It was shown in Chapter 4 to 8 that eigenvectors associated with the data are a major factor affecting the results of SCA methods. The same type of structures but with eigenvectors differing from those of the data sets in Sections 4.3 and 6.2 and with same or different eigenvalues could be investigated. The same general conclusions about SCA methods might be expected but maybe some new features such as the affect of eigenvectors on the SCA methods will be found.

12.5 SCA methods using different restrictions

In Chapter 9, three different restrictions on the directions considered for each simplicity preserving transformation were investigated. All the three restrictions are the same when $k = 0$, and restrictions I and II are always the same when $k = 1$ (Sections 9.2 and 9.3). SCA2 is one of the best SCA methods and the best results generally are obtained by SCA2 when $k = 0$ or 1. So in Chapter 9, only one method, SCA2, using restrictions I and III when $k = 1$ was investigated. It was shown in Sections 9.2 and 9.3 that when $k = 1$, the number of directions available for any individual SCA method using restriction III is more than those using restriction I. However, the results obtained by SCA2 using restriction III are not always more accurate than those using restriction I (for example, simple block structure in Section 9.4.1). The greater number of directions available for restriction III at each step does not guarantee that the maximal improvement in variance using restriction III is more than the maximal improvement in variance using restriction I because sometimes the best maximal improvement in variance might be obtained in a direction available using restriction I but not available using restriction III. Even if the maximal improvement in variance of SCA2 using restriction III at each step is more than that using restriction I, this does not guarantee more accurate results overall (an example of this was given in Section 8.2.1) because this does not guarantee the maximal improvement in variance

overall.

The best results obtained by the SCA methods are defined as the most accurate results among the simple results. When $k = 0$ the results obtained by SCA methods using any restrictions are the same, and they retrieve the simple block and uniform structures. So, if the eigenvectors structure is block structure or uniform structure, just use $k = 0$ and the exact choice of restriction is irrelevant. SCA methods using restriction I obtained the best results for the intermediate structures. So if one restriction from these two restrictions has to be chosen for these data, restriction I is the best choice for the data introduced in Sections 4.3 and 6.2.

If the eigenvectors matrix has an intermediate structure, the best choice of k is 1 (Section 8.5), the best restriction is based on the structure of the eigenvector matrix. Restriction I is the best choice for the intermediate structures introduced in Sections 4.3 and 6.2 because the loadings of the transformation matrix are similar to the loadings of the simple intermediate structures for two dimensional data (Sections 4.7 and 8.5) when $k = 1$. For a given structure, a suitable restriction rather than the number of the directions available at each each iteration is a major factor for SCA methods to get good results.

The analysis in Chapter 9 did not consider all cases. The conclusions obtained by SCA2 can be extended to other variants of the SCA algorithm and applied to other data. In order to compare different restrictions, it is enough to compare the results of SCA2, SCA5 and SCA6 using restrictions I and III when $k = 1$, unless the restrictions are different when $k = 0$.

12.6 Combining approach and hybrid approaches

In order to improve the accuracy of the results obtained by SCA methods, two kinds of approaches were introduced in Chapters 10 and 11. The combined approach combines the best results for each component from different SCA methods using different restrictions and different values of k. These new results are called combined results. The best result of a component is defined as the simple component obtained by all the SCA methods with smallest angle between itself and the corresponding principal component. Overall, the combined results then consist of the best results for each component. By the definition of combined results, the combined results

are necessarily better than the results obtained by any individual SCA method. But in general, the combined results are not orthogonal.

The combined results can come from any SCA method using any restriction and any values of k. But for the examples in Chapter 10, generally, the best result of each component is obtained by SCA2, SCA5 or SCA6 using restriction I when $k = 0$ or 1. This is expected because SCA2, SCA5 and SCA6 are the best SCA methods, and SCA2, SCA5 and SCA6 using restriction I generally get the best results for the data in Sections 1.3, 4.3 and 6.2 when $k = 0$ or 1. For the data used in practice, it is generally enough to consider only the results obtained by SCA2, SCA5 and SCA6 using restrictions I and III and k equal to 0 and 1.

In contrast, the hybrid approaches use different SCA methods with possibly different restrictions and different values of k in different iterations. Two types of hybrid approaches, both using the criterion maximal improvement in variance and $k = 0$ at all iterations were considered. One type of hybrid approach, M^qS with $q \geq 1$ used q multiple updating iterations followed by single updating for all remaining iterations. The other type of hybrid approach S^qM with $q \geq 1$ used single updating for the first q iterations followed by multiple updating for the remaining iterations. As q increases, the first type of hybrid approach, M^qS, is more likely to get similar results to SCA5, whereas the second type of hybrid approach, S^qM is more likely to get similar results to SCA2. This is because the higher is q, the more iterations of M^qS are like SCA5 iterations, and the more iterations of S^qM are like SCA2 iterations. So generally it is enough just to consider the hybrid approaches M^qS and S^qM with $q = 1$ and 2.

In general, the hybrid approaches M^qS and S^qM with $q = 1$ and 2 are at least as accurate as those of SCA2 and SCA5.

12.7 Conclusion and discussion

Simple components can be very good approximations to the corresponding PCs. It is surprising that for all the simple structures the sample simulation results obtained by SCA methods generally are not only easier to interpret but also more accurate than those obtained by PCA. This is because the loadings of the simple components are integers and if the normalizing constant

is omitted for the simple structures, the loadings of all the simple structures are also integers. Furthermore for two dimensional data, the loadings of the transformation matrix $B(b)$ (Section 3.3.2) at every step are more likely to be similar to the loadings of the simple block structure or simple uniform structure (Section 4.7) when $k = 0$. For more than two dimensional data, the multiplication (P in Section 3.3.3) of the transformation matrices at each step is likely to be simple block or simple uniform structures. Similarly, the loadings of transformation matrix using restriction I when $k = 1$ are more likely to be similar to the loadings of the simple intermediate structure in Sections 4.3 and 6.2. The simple components algorithm tends to get the closest simple components to the principal components. If the eigenvectors structures in Sections 4.3 and 6.2 are simple structures, it is possible for the SCA algorithm to retrieve simple structures. In contrast for the sample simulation results, the loadings of the principal components generally are not integers. This makes the sample simulation results obtained by SCA generally more accurate than those obtained by PCA for simple structures. The sample simulation results obtained by SCA are also possibly more accurate than those obtained by PCA for complex structures. For example when $k = 0$ the sample simulation results obtained by SCA for 8 dimensional complex uniform structure are more accurate than those obtained by PCA. This is because SCA2, SCA5 and SCA6 retrieve the simple uniform structure for both simple and complex uniform structures, and the angles between the simple uniform structure and corresponding complex structure are smaller than the difference between the sample simulation results between PCA and this SCA method for simple uniform structure. In other words, the sample simulation results obtained by SCA are more accurate than those obtained by PCA because PCA has no bias but large variance, whereas the angles obtained by SCA have a discrete distribution, SCA has some bias but small variance for complex structures.

For complex structures, the best angles between the simple components and the corresponding principal components generally are obtained if the SCA method gets the corresponding simple structure back exactly. This is because the components of the simple structure is generally the closest simple approximation of the components of the corresponding complex structure by construction. Generally for the complex structures the results of SCA methods are slightly less accurate than those of PCA but are much easier to interpret than those of PCA. So if the

results of PCA are good for some data, the results of SCA should be good.

Generally the results obtained by SCA methods with maximal improvement in variance are at least as accurate but not as simple as those obtained by SCA methods with maximal variance. Further the results obtained by the single SCA method with maximal improvement in variance generally were at least as accurate as and more simple than those obtained by the multiple SCA method with maximal improvement in variance. However, the results obtained by the multiple SCA method with the maximal variance generally were more accurate but not as simple as the results obtained by the single SCA method with the maximal variance. The condition $l_r = l_s$ makes the results obtained by the SCA methods generally less accurate (or the same) but more simple than the results obtained by SCA methods without this condition.

Generally, SCA2, SCA5 and SCA6 are the best SCA methods to use whatever structure and whatever the dimension of the data. Combining the best results of each component of the SCA methods generally improve the results of SCA methods. It was also shown that the results of hybrid approaches M^qS and S^qM with $q = 1$ and 2 when $k = 0$ are possibly more accurate than those obtained by any individual SCA method.

The effect of k and the effect of different patterns of eigenvalues has been investigated. It was shown that eigenvectors associated with the data are a major factor in affecting the results of SCA methods. So the same structures but different eigenvectors from the data sets in Sections 4.3 and 6.2 with the same or different eigenvalues should be investigated in the future.

Three different restrictions to the directions available for the simplicity preserving transformation were discussed. When $k = 0$, all the restrictions are the same and SCA got the simple block and uniform structures back exactly. So whatever the restrictions, SCA got the best results for simple block and simple uniform structures. So, to judge whether a restriction is good or not, it is enough to consider the results of simple intermediate structure when $k = 1$. It was shown that when $k = 1$ restriction I is generally the best choice for the intermediate structures in Sections 4.3 and 6.2. In order to get good results for different intermediate structures, new restrictions can be introduced. For example, by just changing the term 2^k in restriction I to 5^k, 6^k, 7^k etc according to the form of the intermediate structure. In practice, the eigenvectors' structures are a mixture of complex block, uniform and intermediate structures. It is generally

enough to consider the SCA methods using restrictions I and III when $k = 0$ or 1.

In Chapter 9, only SCA2 using restrictions I and III was investigated. The analysis in Chapter 9 can be extended to all the SCA methods when $k = 1$, to the SCA methods using other k, to SCA methods using restrictions I and III for large dimensional data.

Two types of hybrid approaches were considered. The first type of hybrid approach M^qS with $q \geq 1$ used multiple SCA methods in the first q iterations and single SCA methods in the remaining iterations. The second type of hybrid approach S^qM with $q \geq 1$ used single SCA methods in the first q iterations and multiple SCA methods in the remaining iterations. It was shown that a hybrid approach can improve the results of SCA methods. I only investigated the hybrid approaches of single and multiple SCA methods with maximal improvement in variance. The hybrid approaches can be extended to the SCA methods with different criteria. The hybrid approaches can use different SCA methods with different restrictions and different k. Thus, many different hybrid approaches can be tried. In practice, it is enough to consider the hybrid approaches M^qS and S^qM with $q = 1$ and 2 when k is equal to 0.

To sum up, the results obtained by SCA methods can be very good approximation to corresponding PCs whatever the structures of the data, and simple components generally are easier to interpret than PCs. In particular, whatever the data, the sample simulation results of SCA generally are better than PCA for any simple structures. And for large dimensional data, the sample simulation results for complex block and uniform structures when $k = 0$ are better than those obtained by PCA. The sample simulation results for intermediate structures when $k = 1$ are expected to be better than those obtained by PCA. SCA2, SCA5 and SCA6 are the best SCA methods. As the dimension of the data increases, SCA5 and SCA6 are similar. If necessary, combined results and hybrid approaches can make the results of SCA more accurate. However, the results of hybrid approaches are orthogonal but the combined results generally are not orthogonal.

Far from being a narrow technique to enhance the interpretation of PCs, SCA is often compared with the other approaches to enhance the interpretation of PCs in much of the recent research such as Jolliffe and Uddin (2000, 2002), Chipman and Gu (2003), Rousson and Gasser(2004) and Zou, Hastie, and Tibshirani (2004).

Bibliography

[1] Beale, E. M. L., Kendall, M. G. and Mann, D. W. (1967) The discarding of variables in Multivariate Analysis, Biometrika, 54, 357-366.

[2] Bumpus, H. C. (1898) The elimination of the unfit as illustrated by the introduced sparrow, Passer domesticus, Bio. Lectures, Marine Biol. Lab., Woods Hole: 209-226.

[3] Cadima, J. and Jolliffe, I. T. (1995) Loadings and correlations in the interpretation of principal components. J.Appl.Statist, 22, 203-214.

[4] Cadima, J. and Jolliffe, I. T. (2001) variable selection and the interpretation of principal subspaces. J.Agri.Biol.Environ.Statist., 6, 62-79.

[5] Carroll, J. B. (1953) An analytical solution for approximating simple structure in factor analysis. Psychometrika, 18, 23-38.

[6] Carroll, J. B. (1957). Biquartimin criterion for rotation to oblique simple structure in factor analysis. Science, 125, 1114-1115.

[7] Chipman, H. A. and Gu, H. (2003) Interpretation dimension reduction. Submitted for publication.

[8] Euromonitor (1979), European Marketing Data and Statistics, London: Euromonitor Publications, 76-77

[9] Gervini, D. and Rousson, V. (2004), Criteria for Evaluating Dimension-Reducing Components for Multivariate Data, Am. Statistn, 58, 72-76.

[10] Harris, C. W. Kaiser, H. F. (1964) Oblique factor analytic solutions by orthogonal transformations. Psychometrika 29, 347-362.

[11] Hausman, R. E. (1982) Constrained Multivariate Analysis, Studies in the Management Sciences 19.

[12] Hocking, R. R. (1976) The Analysis and Selection of variables in Linear Regression, Biometrics, 32, 1-49.

[13] Hocking, R. R. (1983) Developments in Linear Regression Methodology: 1959-1982, Technometrics, 25, 219-249.

[14] Jackson, J. E. (1991) A User's Guide to Principal Components (1st ed.), New York: Wiley, p4-33; 142-172.

[15] Jeffers, J. N. R. (1967) Two case studies on the application of principal component analysis, Appl.stat.,16,225-236.

[16] Jolliffe, I. T. (1972) Discardings variables in a Principal Component Analysis I: Artificial Data, Applied Statistics Joneal of the Royal Statistical Society, Ser, C.21, 160-163.

[17] Jolliffe, I. T. (1973) Discarding variables in a Principal Component Analysis II: Real Data, Applied Statistics, Journal of the Royal Statistical Society, Ser, C, 22, 21-31.

[18] Jolliffe, I. T. (1987) Rotation of principal components: some comments. J.Climatol., 7, 507-510.

[19] Jolliffe, I. T. (1995) Rotation of principal components: choice of normalization constraints, J.Appl.Statist., 22, 29-35.

[20] Jolliffe, I. T. and Uddin, M. (2000) The Simplified Component Technique: An Alternative to Rotated Principal Components, Journal of Computational and Graphical Statistics, 9, 689-710.

[21] Jolliffe, I. T. and Uddin, M. (2002) A Modified Principal Component Technique Based on the LASSO. J. Comput. Graph. stat. 12: 531-547.

[22] Jolliffe, I. T. (2002a) Multivariate Statistical Methods in Atmospheric Science. Compte-redu de la Iveme Journee Statistique IPSL.

[23] Jolliffe, I. T. (2002b) Principal component Analysis, 2nd edition. Springer, New York.

[24] Jolliffe, I. T., Uddin, M. and Vines, S.K. (2002) Simplified EOFs-three alternatives to rotation. Clim.Res.,20: 271-279.

[25] Kaiser, H. F. (1958) The varimax criterion for analytic rotation in factor analysis, Psychometrika, 23, 187-200.

[26] Manly, B. F. J. (1995) Multivariate statistics methods, Chapman and Hall, London.

[27] McCabe, G. P. (1984) Principal variables, Technometrics. Vol.26, NO.2, 137-145.

[28] Mestas-Nunez, A. M. (2000) Orthogonality properties of rotated empirical modes. Int. J. Climatology, 20, 1509-1516.

[29] Okamoto, M. (1969), "Optimality of Principal Components," in Multivariate Analysis II, ed. P.R. Krishnaiah, New York: Academic Press, 673-685.

[30] Rao, C. R. (1964) The use and interpretation of principal component analysis in applied research, Sankhya, ser. A, 26, 329-358.

[31] Richman, M. B. (1986) Rotation of principal components: a reply. J. Climatol., 6, 293-335.

[32] Richman, M. B. (1987) Rotation of principal components: a reply. J. Climatol., 7, 511-520.

[33] Rousson, V. and Gasser, Th. (2003) Some case studies of simple component analysis, unpublished.

[34] Rousson, V. and Gasser, Th. (2004) Simple Component Analysis, Applied Statistics, 53, 539-555.

[35] Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. J. Royal Statist. Soc. B, 58, 267-288.

[36] Thompson, M. O., Vines, S. K. and Harrington, K. (1999) Assessment of bloodvolume flow in the uterine artery; the influence of arterial distensibility and waveform abnormality. Ultsound obstet. Gyn., 14, suppl. 1, 71.

[37] Vines, S. K. (2000) Simple principal component, Applied Statistics, 49, 441-451.

[38] Zou, H. and Hastie, T. (2003), Regression shrinkage and selection via the elastic net, with applications to microarrays, Technical report, Statistics Dept. Stanford University.

[39] Zou, H., Hastie, T. and Tibshirani (2004) Sparse Principal Component Analysis, Technical report, Statistics Dept. Stanford University.